

Celebrating 50 Years of Traffic Flow Theory

DETAILS

510 pages | 8.5 x 11 | PAPERBACK

ISBN 978-0-309-43245-0 | DOI 10.17226/22095

AUTHORS

Multiple Authors; Technical Activities Division; Transportation Research Board; National Academies of Sciences, Engineering, and Medicine

BUY THIS BOOK

FIND RELATED TITLES

Visit the National Academies Press at NAP.edu and login or register to get:

- Access to free PDF downloads of thousands of scientific reports
- 10% off the price of print titles
- Email or social media notifications of new titles related to your interests
- Special offers and discounts



Distribution, posting, or copying of this PDF is strictly prohibited without written permission of the National Academies Press. (Request Permission) Unless otherwise indicated, all materials in this PDF are copyrighted by the National Academy of Sciences.

TRANSPORTATION RESEARCH
CIRCULAR

Number E-C197

September 2015

**Celebrating 50 Years of
Traffic Flow Theory**

A Symposium

August 11–13, 2014
Portland, Oregon

TRANSPORTATION RESEARCH BOARD

**TRANSPORTATION RESEARCH BOARD
2015 EXECUTIVE COMMITTEE OFFICERS**

Chair: Daniel Sperling, Professor of Civil Engineering and Environmental Science and Policy;
Director, Institute of Transportation Studies, University of California, Davis

Vice Chair: James M. Crites, Executive Vice President of Operations, Dallas–Fort Worth
International Airport, Texas

Division Chair for NRC Oversight: Susan Hanson, Distinguished University Professor
Emerita, School of Geography, Clark University, Worcester, Massachusetts

Executive Director: Neil J. Pedersen, Transportation Research Board

**TRANSPORTATION RESEARCH BOARD
2014–2015 TECHNICAL ACTIVITIES COUNCIL**

Chair: Daniel S. Turner, Emeritus Professor of Civil Engineering, University of Alabama,
Tuscaloosa

Technical Activities Director: Ann M. Brach, Transportation Research Board

Peter M. Briglia, Jr., Consultant, Seattle, Washington, *Operations and Preservation Group
Chair*

Alison Jane Conway, Assistant Professor, Department of Civil Engineering, City College of
New York, New York, *Young Members Council Chair*

Mary Ellen Eagan, President and CEO, Harris Miller Miller and Hanson, Inc., Burlington,
Massachusetts, *Aviation Group Chair*

Barbara A. Ivanov, Director, Freight Systems, Washington State Department of Transportation,
Olympia, *Freight Systems Group Chair*

Paul P. Jovanis, Professor, Pennsylvania State University, University Park, *Safety and Systems
Users Group Chair*

D. Lane, Associate Principal Research Scientist, Virginia Center for Transportation Innovation
and Research, *Design and Construction Group Chair*

Hyun-A C. Park, President, Spy Pond Partners, LLC, Arlington, Massachusetts, *Policy and
Organization Group Chair*

Harold R. (Skip) Paul, Director, Louisiana Transportation Research Center, Louisiana
Department of Transportation and Development, Baton Rouge, *State DOT Representative*

Ram M. Pendyala, Frederick R. Dickerson Chair and Professor of Transportation, Georgia
Institute of Technology, *Planning and Environment Group Chair*

Stephen M. Popkin, Director, Safety Management and Human Factors, Office of the Assistant
Secretary of Transportation for Research and Technology, Volpe National Transportation
Systems Center, Cambridge, Massachusetts, *Rail Group Chair*

Robert Shea, Senior Deputy Chief Counsel, Pennsylvania Department of Transportation, *Legal
Resources Group Chair*

Eric Shen, Director of Transportation Planning, Port of Long Beach, *Marine Group Chair*

David C. Wilcock, Vice President and National Practice Leader for Rail and Transit, Michael
Baker, Jr., Inc., Norwood, Massachusetts, *Public Transportation Group Chair*

TRANSPORTATION RESEARCH CIRCULAR E-C197

Celebrating 50 Years of Traffic Flow Theory

A Symposium

Standing Committee on Traffic Flow Theory
and Characteristics Summer Meeting

August 11–13, 2014
Portland, Oregon

September 2015

Transportation Research Board
500 Fifth Street, NW
Washington, D.C.
www.TRB.org

TRANSPORTATION RESEARCH CIRCULAR E-C197

The **Transportation Research Board** is one of seven programs of the National Academies of Sciences, Engineering, and Medicine. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal.

The **Transportation Research Board** is distributing this E-Circular to make the information contained herein available for use by individual practitioners in state and local transportation agencies, researchers in academic institutions, and other members of the transportation research community. The information in this circular was taken directly from the submission of the authors. This document is not a report of the National Academies of Sciences, Engineering, and Medicine.

Symposium Organizing Committee

Robert Bertini, *California Polytechnic State University, San Luis Obispo, Chair*
Soyoung Ahn, *University of Wisconsin*
Tegan Enloe, *DKS Associates*
Vikash Gayah, *The Pennsylvania State University*
Nikolas Geroliminis, *École Polytechnique Fédérale de Lausanne*
Victor Knoop, *Delft University of Technology*
Ludovic Leclercq, *Université de Lyon, IFSTTAR/ENTPE*
Yanfeng Ouyang, *University of Illinois at Urbana–Champaign*
Haizhong Wang, *Oregon State University*
Miranda Wells, *HDR*

Standing Committee on Traffic Flow Theory and Characteristics

Robert Bertini, *Chair*
Soyoung Ahn, *Secretary*

Constantinos Antoniou
Christine Buisson
KooHong Chung
Winnie Daamen
Soumya Dey
Jing Dong
Nathan Gartner
Vikash Gayah
Nikolas Geroliminis
Eric Gonzales
Mohammed Hadi
Samer Hamdar

Victor Knoop
Jorge Laval
Ludovic Leclercq
George List
Hani Mahmassani
Michael Mahut
Stephen Mattingly
Monica Menendez
Yu Nie
Yanfeng Ouyang
Vincenzo Punzo
Meead Saberi Kalae

Majid Sarvi
Robert Sheehan
Alexander Skabardonis
James Sturrock
Tomer Toledo
Avinash Unnikrishnan
Hans Van Lint
Peter Vortisch
Yibing Wang
Zhongren Wang
Ethan Xuan
Marguerite Zarrillo
H. Zhang

Transportation Research Board

Richard A. Cunard, *Senior Program Officer*
Freda R. Morgan, *Senior Program Associate*

Transportation Research Board
500 Fifth Street NW
Washington, D.C.
www.TRB.org

Preface

As one of the long-time international members of the Traffic Flow Theory and Characteristics Committee and a long-time head of its German counterpart, I want to express my sincere congratulations on 50 years of traffic flow theory.

I remember the ancestors of this committee—like Bob Herman, a former researcher at GM, who started investigations of car-following behavior by distance measurements of a vehicle in a platoon with a thread wound on a coil and connected with the preceding vehicle. Also notable was the analysis of “Investigation of Traffic Dynamics by Aerial Photogrammetry Techniques” by Joseph Treiterer and Jeffrey A. Myers from Ohio State University. All this pioneering work of traffic flow description and traffic pattern formation was based on the ideas of Bruce Greenshields, who first developed the terms “traffic density,” “traffic flow,” and “(mean) speed;” their definitions; and the corresponding measuring instructions more than 80 years ago. It was therefore an utmost concern to celebrate Greenshields’s legacy, which the committee honored with a midyear meeting in 2008, in Woods Hole, Massachusetts, commemorating 75 years of Greenshields’s famous publication. The resounding success started a series of meetings: the 2008 symposium at Woods Hole; “Does Traffic Data Support Traffic Models?” in Annecy, France, in 2010; the Symposium on Innovations in Traffic Flow Theory and Highway Capacity and Quality of Service in Fort Lauderdale, Florida, in 2012; and now, in 2014, the committee’s 50th anniversary celebration in Portland, Oregon.

The observation during my time working with the committee is the trend away from macroscopic modeling and toward microscopic simulation because of the enormously increased power capacity of computers. Another trend is based on the influence of modern communication technology and driver-assistance systems on traffic pattern formation. The applications of short-range communication technology connecting vehicles with each other and with infrastructure facilities enables new traffic control possibilities and opens new challenges for traffic flow description and modeling. Cooperative driving with data exchange governing braking actions, crash sensor data, steering maneuvers, windscreen wiper positioning, and warning light positioning will shape the traffic of the future and redesign the requirements of traffic flow theory. The airbag sensor technology will provide us with high-quality approach data.

Does traffic flow theory offer answers to the way in which rapid and safe data transfer has to be organized with prioritization of surrounding vehicles?

It is already clear that the active distance warning equipment will enhance the knowledge of distance and speed and give support for measurements in the “vehicle-to-vehicle” context.

From “vehicle-to-infrastructure” communication, such as automatic tolling and roadside warning systems, destination data can be derived (vehicle probe or floating car data). These data find their way into traffic signal preemption and operating condition checks of computer-aided operating control systems for transit, an alternative to floating car data via mobile telephone. Traffic flow theory can help in the design of safe data transfer including vehicle classification and identification, as well as for the design of displays showing hazardous conditions based on multifactorial interpretation of actively transmitted vehicle data. The multifactorial analysis links the data of the vehicle to infrastructure communication with wheel revolution counter and odometer data.

One of the accomplishments of the committee is the initiation of the “Traffic Flow Monograph,” edited by the Federal Highway Administration. This basic compendium is a terrific

support for academia and practitioners, even in its present form, to answer questions about the simulation of traffic systems, emissions reduction through better traffic management, empirical evaluation based upon on-road measurements, and incident management in intelligent transportation. It will be exciting to look forward to a new edition that addresses some of the challenges explained above.

The stimulations to the *Highway Capacity Manual* are not to be underestimated. Sharing the ideas with representatives of well-respected universities from the United States and Canada always enriches the committee and so I wish the committee a bright future and many more years of inspiring work—*ad multos annos!*

—Reinhart Kühne, Germany

Contents

FUNDAMENTAL DIAGRAM

Capacity Drop: A Comparison Between Stop-and-Go Wave and Queue Congestion at Lane-Drop Bottleneck	3
<i>K. Yuan, V. Knoop, L. Leclercq, and S. Hoogendoorn</i>	
An Overview of “Revisiting the Empirical Fundamental Relationship”	18
<i>B. Coifman</i>	
Universalities in Fundamental Diagrams of Cars, Bicycles, and Pedestrians	25
<i>A. Seyfried, E. Andresen, M. Boltes, St. Holl, W. Mehner, A. Schadschneider, and J. Zhang</i>	
Human Factors in the Fundamental Diagram	30
<i>D. Ni, L. Li, H. Wang, and C. Jia</i>	

TRAFFIC FLOW MODELS

Congestion Scenario-Based Vehicle Classification Detection Models Based on Traffic Flow Characteristics and Observed Event Data	49
<i>H. Wei, Q. Ai, H. Liu, Z. Li, and H. Wang</i>	
Optimal Velocity Model with Dual Boundary Optimal Velocity Function	65
<i>H. Wang</i>	
Data Fusion Solutions to Compute Performance Measures for Urban Arterials	82
<i>H. van Lint, R. Bertini, and S. Hoogendoorn</i>	
Modeling Acceleration Behavior in a Connected Environment	87
<i>A. Talebpour and H. Mahmassani</i>	

TRAFFIC CONTROL

Traffic Flow Theory Milestones in Developing the TEXAS Model for Intersection Traffic in the Early 1970s	95
<i>T. Rioux</i>	
Variable Speed Limit Control to Increase Discharge Rates of Freeway Incident Bottlenecks	117
<i>D. Chen and S. Ahn</i>	
A Real-Time Signal Control Strategy for Mitigating the Impact of Bus Stops on Urban Arterials	133
<i>C. Chavis and E. Christofa</i>	

FREEWAY TRAFFIC ANALYSIS

Influential Subspaces of Connected Vehicles in Highway Traffic	151
<i>K. Jerath, V. Gayah, and S. Brennan</i>	
The Heterogeneity of Capacity Distribution Among Different Freeway Lanes	161
<i>K. Xie, K. Ozbay, and H. Yang</i>	

CALIBRATION

Validating the Cost-Effectiveness Model for California’s Freeway Incident Management Program.....179
M. Mauch, A. Skabardonis, and L. Davies

Calibrating Multilane First-Order Traffic Flow Model with Endogenous Representation of Lane-Flow Equilibrium.....192
Y. Shiomi and T. Kozono

Heterogeneous Nonlinear Car-Following Laws for Traffic Oscillation Prediction.....210
C. Rhoades, X. Wang, and Y. Ouyang

Using Big Data and Efficient Methods to Capture Stochasticity for Calibration of Macroscopic Traffic Simulation Models.....215
S. Mudigonda and K. Ozbay

EMPIRICAL OBSERVATIONS OF TRAFFIC ANALYSIS

Continuous Flow Metering Alternative Solution to Alleviate Congestion on Interstate 70 Eisenhower Johnson Memorial Tunnel.....233
S. Marlina, B. Janson, and S. Sobhi

Measuring the Safety Impact of Road Infrastructure Systems on Driver Behavior: Vehicle Instrumentation and Exploratory Analysis.....249
J. Schorr, S. Hamdar, and C. Silverstein

NETWORKWIDE MODELING AND CONTROL

Toward a Systematic Exploration of the Influence of Route Choices on a Network Level of Performance.....267
C. Parzani, L. Leclercq, N. Benoumechiara, and D. Villegas

Macroscopic Relationship Between Networkwide Traffic Emissions and Fundamental Properties of the Network.....284
R. Shabihkhani and E. Gonzales

Existence, Stability, and Mitigation of Gridlock in Beltway Networks.....300
W. Jin

Effects of Segregating Buses and Cars in a Congested, Non-Steady State Street Network.....310
N. Saade, W. Gu, and M. Cassidy

Stochastic Approximations for the Macroscopic Fundamental Diagram.....315
J. Laval, F. Castrill’on, and Y. Zhou

POSTER PRESENTATIONS

Adapting Car Traffic Models and Concepts to Bicycle Traffic.....321
A. Manar and G. Cao

Gaussian Approximation for Modeling Traffic Flow on a Homogeneous Road Segment.....334
B. Vachta, X. Qin, and J. Kimn

Investigation of Performance and Lane Utilization Within a Passing Lane on a Two-Lane Rural Highway.....350
Z. Freedman and A. Al-Kaisy

Urban Road Network Macroscopic Fundamental Diagram Analysis Under Vehicular Ad-Hoc Networks Environment	361
<i>Z. Xu, P. Jin, and B. Ran</i>	
Real-Time Control of Queue Spillbacks on Signalized Arterials	364
<i>M. Ramezani, N. de Lamberterie, A. Skabardonis, and N. Geroliminis</i>	
Macroscopic Evaluation of Incident-Induced Driver Behavior Changes	369
<i>Z. Rahman and S. Mattingly</i>	
Current Validation of Traffic Models: Are They Correct for Uncertainty Analysis? [Abstract Only]	385
<i>J. Casas, V. Punzo, J. Perarnau, and M. Montanino</i>	
A Time Series Analysis of Highway Capacity: Case Study of Georgia SR-400	386
<i>S. Dong, H. Wang, and J. Li</i>	
Collaborative Merging Behaviors and Their Impacts on Freeway Ramp Operations Under Connected Vehicle Environment	392
<i>Y. Xie, H. Zhang, N. Gartner, and T. Arsava</i>	
A Second-Order Lagrangian Macroscopic Traffic Flow Model for Freeways	409
<i>Z. Zhou and P. Mirchandani</i>	
 APPENDIXES	
A. Keynote Speakers’ Powerpoint Presentations	
Reflections on 50 Years of the TRB Committee on Traffic Flow Theory and Characteristics	427
<i>Nathan Gartner</i>	
Exploring the Impact of Microscopic Features of Traffic on Macroscopic Patterns	439
<i>Vincenzo Punzo</i>	
Traffic Flow Theory in the Era of Autonomous Vehicles	447
<i>Michael Zhang</i>	
Challenges in Pedestrian Flow Modeling	451
<i>Serge Hoogendoorn</i>	
Trajectories in 3-D: Unifying Model Calibration and Network Performance Analysis	458
<i>Hani S. Mahmassani</i>	
 B. PAPER PRESENTATIONS	
Automatic Fitting Procedure for the Fundamental Diagram	473
<i>V. Knoop and W. Daamen</i>	
Accuracy of Pedestrian and Traffic Flow Models: Meaningful Quantifications	491
<i>F. van Wageningen-Kessels, W. Daamen, and S. Hoogendoorn</i>	
The Effect of Stochastic Volatility in Predicting Highway Breakdowns	509
<i>P. Ossensbruggen and E. Laflamme</i>	

Fundamental Diagram

FUNDAMENTAL DIAGRAM

Capacity Drop
*A Comparison Between Stop-and-Go Wave
 and Queue Congestion at Lane-Drop Bottleneck*

KAI YUAN

VICTOR L. KNOOP

SERGE P. HOOGENDOORN

Delft University of Technology, the Netherlands

LUDOVIC LECLERCQ

Université de Lyon, France

In freeways, the maximum traffic flow through a bottleneck is usually higher than the outflow of congestion there. This phenomenon is called the capacity drop. In literature, there are considerable debates about the mechanism causing this phenomenon. This paper studies the mechanism by analyzing real-life data of 2 different days. The traffic states downstream of a lane drop are analyzed. It is observed that the outflow of a stop-and-go wave on the three-lane section is lower than that of a standing queue upstream from the lane-drop bottleneck. A more detailed analysis shows the phenomenon on lane level. Finally, data from 2 days show a common feature on flow distribution over lanes. This finding shows that even in congestion states, the flow in shoulder lane (slow lane) can be lower than that in other lanes in the three-lane section because of lower density. Moreover, it is found that close to the bottleneck, a larger part of the flow is in the median lane. After several hundred meters the lane flow distribution normalizes to equilibrium, indicating much lane changing out of the median lane directly downstream of the lane-drop bottleneck. At a four-lane section upstream from the bottleneck, a large number of lane changes occur. The understanding of the mechanism behind the capacity drop, as well as the sizes of the capacity drop, might lead to measures to reduce delay. Moreover, the flow distribution can contribute to lane-changing models closely resembling reality.

INTRODUCTION

Generally congestion can be divided into two classes: stop-and-go waves, propagating parts of congestion with two fronts moving upstream along a freeway, and standing queue, with its head fixed at a bottleneck. An active bottleneck is a bottleneck with a free-flow situation downstream and a traffic jam upstream. The activation of a bottleneck signals the onset of a standing queue. Theoretically downstream of an active bottleneck the outflow of the standing queue should be the maximum flow on the road, or capacity. However, the queue discharging rate of congestion is often lower than the maximum flow on a road without congestion. This phenomenon is called the bottleneck capacity drop.

Researchers have observed the capacity drop phenomenon for decades at bottlenecks. Those observations point out that the range of capacity drop, the difference between the bottleneck capacity and the queue discharging rate, can vary in a wide range. The capacity of the

road and the queue discharging flow are essential for the total delay on the road. Hall and Agyemang-Duah (1) report a drop of around 6% on empirical data analysis. Cassidy and Bertini (2) place the drop ranging from 8% to 10%. Srivastava and Geroliminis (3) observe that the capacity falls by approximately 15% at an on-ramp bottleneck. Chung et al. (4) present a few empirical observations of capacity drop from 3% to 18% at three active bottlenecks. They show at the same location the capacity drop can range from 8% to 18% (excluding the influences of light rain). Cassidy and Rudjanakanoknad (5) observe capacity drop ranging from 8.3% to 14.7%. Oh and Yeo (6) collect empirical observations of capacity drop in nearly all previous research before 2008. The drop ranges from 3% to 18%.

Even though a large amount of research effort has gone into the capacity drop, some significant the macroscopic features on capacity drop are still unclear. For example, it is not clear to what extent the capacity is reduced when different congestion occurs upstream. Moreover, the amount of traffic on each lane (flow distribution over lanes) is unclear, especially at the downstream of a bottleneck with compulsory merging behaviors upstream. Hence, this paper tries to show more empirical observations to forward traffic research to reveal more empirical features. These findings can contribute to a better understanding of the traffic processes, possibly leading to control principles mitigating congestion. The paper also gives an indication of the lane change behavior at the bottleneck locations.

This paper answers the following question: What are differences between traffic states downstream of stop-and-go waves compared to downstream of standing queues at the same site? In answering this question, researchers use the following four subquestions. First, to what extent does the capacity reduce downstream of a stop-and-go wave? Most of the previous research observes capacity drop phenomenon at active bottlenecks. Few of those studies (7) reveal features of capacity drop downstream of a stop-and-go wave. This study presents empirical observations of capacity drop in a stop-and-go wave. Second, to what extent does the outflow of congestion (i.e., the capacity with congestion upstream) vary at the same road section without other disturbances such as weather and road layouts? In short, this subquestion discusses the stochasticity of the outflow of the queue. Previous research shows that discharging flows of standing queues at one bottleneck only exhibit small deviations (2); but that research only targets standing queue at an active bottleneck. In contrast to the standing queue, where traffic states are limited in a narrow range because the road layout dictates the congested traffic state upstream, different stop-and-go waves can result in different congestion states, including standing queues and stop-and-go waves. The study of stop-and-go waves can enlarge the observation samples. Third, what is the flow in each lane in queue discharge conditions? This might shed light on the capacity drop as well. Fourth, what is the traffic flow distribution over lanes downstream of a bottleneck with compulsory merging behaviors upstream, especially locations near bottlenecks?

To answer these four subquestions, this paper studies a traffic scenario where a standing queue forms immediately after a stop-and-go wave passes. It seems that the standing queue is induced by the stop-and-go wave. In this scenario, there can be at least two congestion states and two outflow states observed at the same road section on the same day.

The remainder of the paper is set up as follows. Section 2 describes methodologies applied in this paper. This section applies shockwave analysis to recognize those different congestion. Section 3 shows the study site and the study data. In Section 4, empirical observations are presented, including various traffic states and flow distribution in each lane. Finally, Section 5 presents the conclusions.

METHODOLOGY

This paper targets a homogeneous freeway section with a lane-drop bottleneck upstream. In the expected scenario, a standing queue forms immediately after the passing of a stop-and-go wave. It seems like the bottleneck is activated by the stop-and-go wave. In this way, researchers can compare the outflows of congestion at that location, and possible location-specific influences are excluded from the analysis.

Since the differences in the capacity drop (in standing queues) between any 2 days at the same bottleneck are in a small range among days (2), it is difficult to observe standing queues in distinctly different congestion states at the same bottleneck. However, the congestion level in a stop-and-go wave is considerably different from the congestion in a standing queue. Congestion level is represented by vehicle speed in the congestion and density. Previous research (4, 8) shows that the capacity drop is strongly related to the congestion level, hence, it is expected that downstream of a stop-and-go wave traffic differs from downstream of a standing queue. This way, by combining several state points at the same road stretch, can be observed empirically, including free flow and congestion states. Shockwave analysis is applied to identify those congestion states qualitatively.

By comparing the outflows downstream of congestion, this paper shows the capacity drop corresponding to the two different congestion types: stop-and-go wave and standing queue. The key to the traffic state analysis is to identify those traffic states. To avoid unnecessary deviations, this paper applies slanted cumulative counts to calculate flow. Both of these two outflows are flow detected downstream of the congestion. There are repetitive observations, for the duration of congestion until the congestion is dissolved, and there are no other influences from downstream apart from the stop-and-go wave. The outflow of a stop-and-go wave can be detected at some location where the speed returns to the free-flow speed after the break down phenomenon, and the discharging flow can be detected at each location downstream of an active bottleneck.

The states that occur are determined using shockwave analysis. Figure 1 shows the resulting traffic states, including the regions in space-time where the outflows can be measured. For the sake of simplicity, the authors chose triangular fundamental diagrams. Figure 1a shows these fundamental diagrams, the smaller one for the three-lane section and the larger one for the four-lane section. The outflow of a stop-and-go wave, shown as State 5, and discharging flow of standing queue, shown as State 6, both lie in the free-flow branch, see Figure 1. The flow in both of these two states is lower than the capacity shown as State 1 to represent the capacity drop. A stop-and-go wave, State 2 in Figure 1, propagates upstream to the bottleneck, and this triggers a standing queue, State 4. Figure 1b shows that once the bottleneck has been activated both of States 5 and 6 can be observed in the downstream of the bottleneck. The further away from the bottleneck, the longer time State 5 can be observed. Note that because States 5 and 6 are always located in the free-flow branch, the shockwave between these two states is always a positive line parallel to the free-flow branch. Therefore, in Figure 1b the shockwave between States 5 and 6 is always the same in $x - t$ plot, no matter which state shows a higher flow.

Hence, for measuring the outflow observations at locations far away from the bottleneck are preferred. In that case, the outflow of a stop-and-go wave can be measured for a long enough time and compared clearly.

With the same methodology, different outflow features in different lanes are analyzed. This shows the performance of each lane during the transition from outflow of stop-and-go wave to queue discharging flow. This paper applies slanted cumulative counts to calculate the outflow

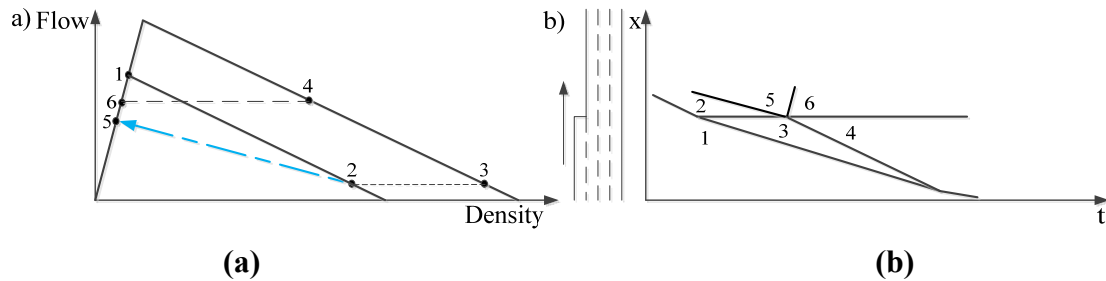


FIGURE 1 Shockwave analysis on one traffic scenario at a lane-drop bottleneck: (a) fundamental diagram and (b) $x-t$ plot.

in each lane. Note that in the Netherlands the rule is “Keep Right Unless Overtaking.” This asymmetric rule might lead to a different lane choice, for instance for slugs and rabbits (9), as well as leading to different traffic operations.

DATA HANDLING

This paper reveals the flow distribution in each lane as a function of average density over lanes in section 4.4. The density (ρ), which is estimated through dividing flow (q) by space-mean speed v_s , is necessary.

In the Netherlands, loop detector data is time mean speed (v_T) and flow (q). Knoop et al. (10) point out the substantial difference between the time-mean speed v_T and space-mean speed v_s , especially when the speed is in congestion. Yuan et al. (11) present a correction algorithm based on flow-density relations to calculate space-mean speed. This method requires that traffic states should be on the linear congested branch of the fundamental diagram. However, this paper considers acceleration states downstream a bottleneck, so the authors need another method. Knoop et al. (10) show an empirical relation between space-mean speed and time-mean speed (Figure 2). The space-mean speed actually is estimated as harmonic speed. This relation is applied to space-mean speed calculation in Ou (12). This paper also applies the relation to calculate the space-mean speed and the density.

DATA

The data analyzed are 1-min aggregated, collected around a lane-drop bottleneck on the freeway A4 in the Netherlands. This paper considers the northbound direction just around Exit 8 (The Hague) in A4, shown in Figure 3. The layout of the study site is shown in the right part of Figure 3. The targeted bottleneck is a lane-drop bottleneck, which is circled in Figure 3. Downstream of this bottleneck, there is another lane-drop bottleneck next to Exit 7. Drivers in the targeted road section are driving from a four-lane section to a three-lane section (the upward direction in Figure 3), so a lane-drop bottleneck occurs. The data are collected from 10 locations with approximately 500-m spacing between them, giving a total length of around 5 km. There are two detectors in the four-lane section, followed by eight in the three-lane section. This paper does not consider detectors further downstream, because vehicles will change into the shoulder

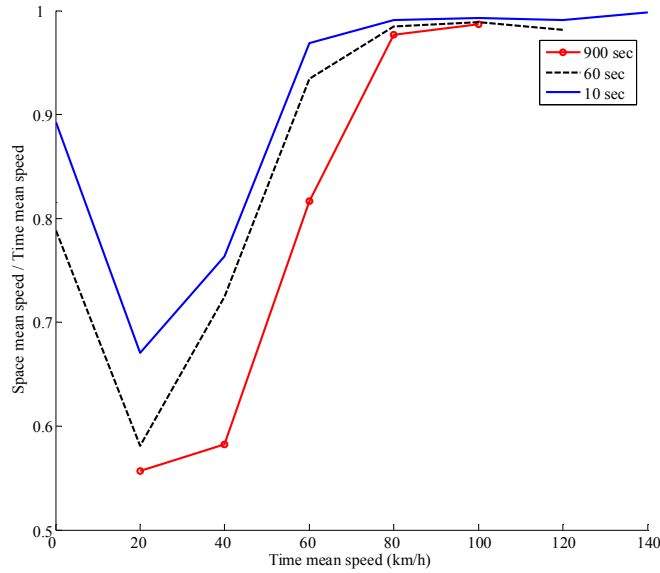


FIGURE 2 The impact of difference between time-mean speed and harmonic mean speed: 10-s aggregation (blue line), 60-s aggregation (black dashed line), and 900-s aggregation (red line with circles). [From Knoop et al. (10)].

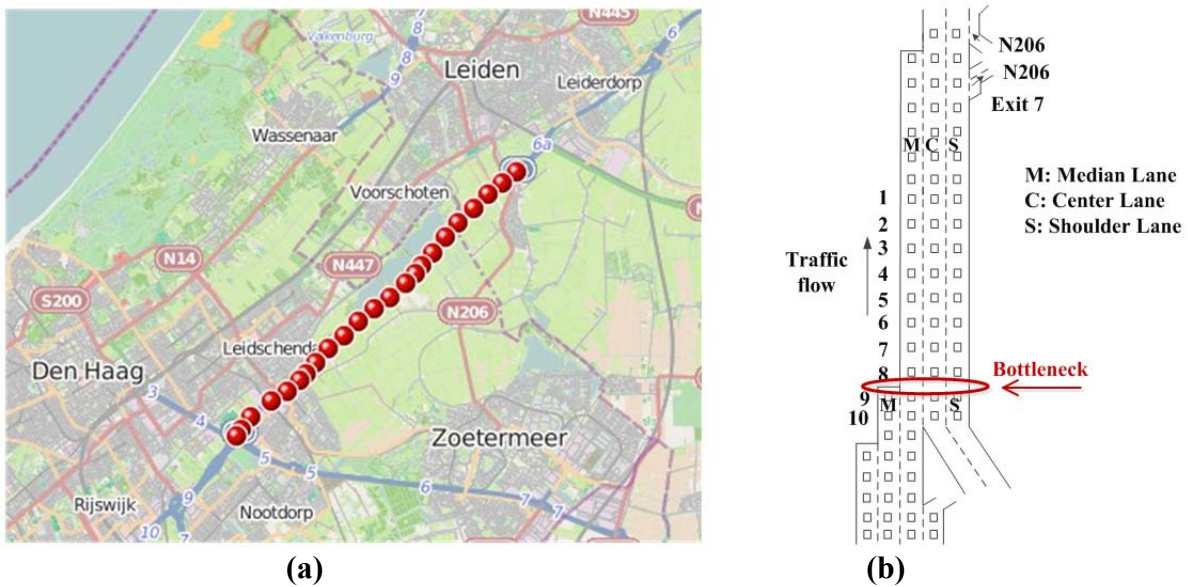


FIGURE 3 Open-street figure of (a) targeted section in freeway A4 in the Netherlands shown in red dots, and (b) the layout of the study site. The bottleneck is a lane-drop bottleneck highlighted with a red circle. This paper only targets 10 locations. The total distance from Location 1 to Location 10 in the freeway is approximately 4.5 km. The bottleneck is around 6.5 km from the downstream off-ramp.

lane to leave the freeway through Exit 7, possibly leading to external disturbances, for instance lane-changing near the off-ramp.

Data for analysis are collected on 2 days: Monday, May 18, 2009, and Thursday, May 28, 2009. Figure 4 shows the speed contour plots in the study section on 2 days. There are two similar traffic situations on both days. The first event is a stop-and-go wave. On May 18 the stop-and-go wave originated from the lane-drop bottleneck near Exit 7 at about 16:45. On May 28 the stop-and-go wave entered the selected stretch from further downstream at around 16:55. At 17:40 and 17:50 (May 18 and 28, respectively), the next stop-and-go wave reached the lane-drop bottleneck. Downstream of the second stop-and-go wave there is congestion. In order to avoid influences of this congestion, this study ends the analysis before the entering of the second stop-and-go wave when calculating the outflows. When analyzing the flow distribution, the data from 16:00 to 19:00 are analyzed. During the targeted period, there is no other influence from downstream (i.e., the bottleneck is active).

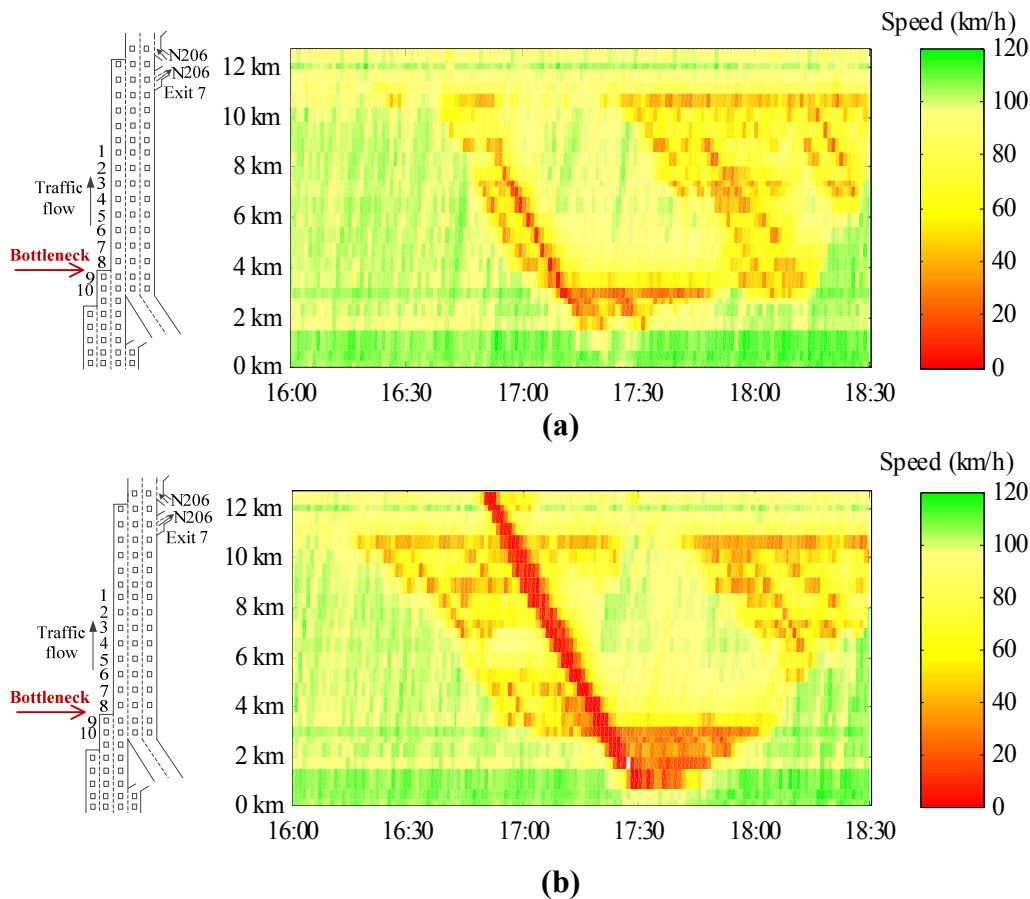


FIGURE 4 Layout of the study site and data on two days for study: (a) speed contour on May 18, 2009, and (b) speed contour on May 28, 2009. The lane-drop bottleneck located between Detectors 8 and 9 is activated by a stop-and-go wave from downstream. The numbers show locations of detectors. This study is restricted to 10 locations around the targeted lane-drop bottleneck.

RESULTS

This section first presents the different states and then the capacity estimates; subsequently, the lane-specific features are discussed.

State Identification

This section describes empirical observations. Generally, the empirical observations are in line with the expectations presented in Section 2. The outflow of the stop-and-go wave and the discharging flow of the standing queue are clearly distinguishable. Figure 4 shows empirical slanted cumulative counts across three lanes at eight locations downstream of the bottleneck on 2 study days. The arrow in each figure shows the shockwave, which propagates downstream from the bottleneck. This means the traffic is in a free-flow state and not influenced by the off-ramp downstream.

This shockwave separates the outflow of stop-and-go wave from the discharging flow of standing queue. This shockwave has been expected in Section 2 (see Figure 1*b*). At one location, the authors first observe the outflow of the stop-and-go wave and then observe the discharging flow of the standing queue. First, they find the outflow of the stop-and-go wave only directly downstream of the stop-and-go wave. The wave travels upstream, from Location 1 to Location 8. Once it reaches Location 8, the traffic state changes, with a wave propagating downstream, which takes some time before it reaches Location 8. During that whole time at Location 1 the outflow of the stop-and-go wave can be detected.

The discharging flows found for the 2 days are constant for each day, at 6,040 veh/h (May 18) and 5,700 veh/h (May 28), see Figure 5. Although they are different for both days, the flows are remarkably constant over time. There is also a difference between the flows downstream of the standing queues on May 18 and 28. This holds for all locations downstream of the bottleneck, including the acceleration phase. The flow is different but constant for both days. During the acceleration process, the density continuously decreases. Since the flows differ for the 2 days, the speeds must differ for the 2 days for situations with an equal density. This

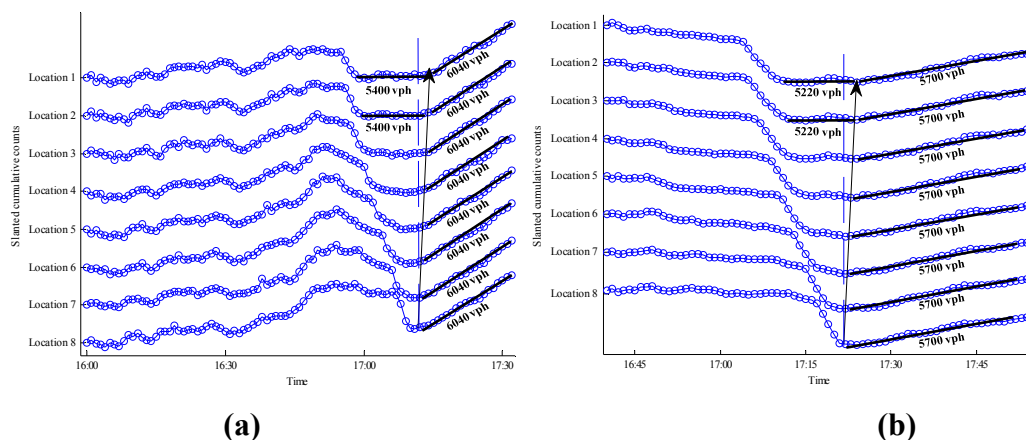


FIGURE 5 Slanted cumulative counts across three lanes at eight locations downstream of the bottleneck on 2 days: (a) May 18, 2009, and (b) May 28, 2009.

means that drivers leave a larger gap than necessary in the day with the lower flow (May 28), since apparently, given the speed–density relationship for the other day, they can drive with higher speeds given the spacing.

Moreover, the downstream direction of the shockwave implies that the off-ramp (Exit 7 in Figure 3) does not influence the discharging flow. Oh and Yeo (6) imply that the off-ramp at the downstream location mitigates the capacity drop. At the study site, the off-ramp that is located far away has no effects. The shockwaves propagating downstream imply no influence from downstream.

Capacity Estimation

Figure 6 shows the capacities (with standing congestion upstream) that are the outflow of congestion at a homogeneous three-lane freeway section. In Figure 6, all red dashed lines show the slanted cumulative curves at the downstream locations, and the blue bold lines represent speed evolution there. All parts in Figure 6 show firstly a decrease of flow (during the time the stop-and-go wave is present), indicated by a cumulative flow line with a negative slope. Afterwards, at Location 1 the flow is constant for about 20 minutes, at approximately 5,400 veh/h on May 18 and 5,220 veh/h on May 28. Figures 6c and d show the slanted cumulative curves for Location 8, just downstream of the bottleneck. After the stop-and-go wave reaches Location 8, the jam soon transforms into a standing queue, and the outflow increases up to 6,040 veh/h and 5,700 veh/h, respectively. These two discharging flows propagate downstream from the bottleneck and reach Location 1. The higher outflow (6,040 and 5,700 veh/h) is not temporary and remains for at least 15 min at each location. The solid black line in each of the figures indicates a flow to which the slanted cumulative curve can be compared. In each figure, the increasing slope of black lines shows that the outflow of stop-and-go wave is lower than the discharging flow of the standing queue. Typically, it is found that the outflow of the stop-and-go wave lies in the range of 5,220 to 5,400 veh/h, and the outflow of the standing queue is in the range of 5,700 to 6,040 veh/h. All data points are collected in Table 1. The number of states corresponds to those in Figure 1.

States 2, 4, 5, and 6 in Figure 1a are identified quantitatively. States 2 and 4 stand for congestion states. States 5 and 6 represent states of capacities. Thus, a correlation between the type of congestion and its outflow is found. In fact, the outflow of a stop-and-go wave is lower than the outflow of a standing queue at the same location.

Outflows in Each Lane

When congestion occurs each lane presents different features regarding to outflows. In Figure 7, slanted cumulative counts and speed in each lane are presented, shown as a red dashed line and a blue bold line, respectively. Slow vehicles and trucks usually drive in the shoulder lane because of the Keep Right Unless Overtaking rule. Therefore, the flow and speed detected in each lane at the same location differ from each other. In parts a and b of Figure 6, aggregated data over three lanes show an increase of outflow at the moment the wave separating the outflow from the stop-and-go wave and the outflow from the standing queue reaches the detector. In parts a and c of Figure 7, this increase of the outflow is observed in the median and center lane at Location 1 on May 18, 2009, but not in the shoulder lane. On May 28 this increase is found in all lanes. The lack of change in flow in the shoulder lane is remarkable, giving the possibility that the difference in the capacity drop is a result of the vehicle composition.

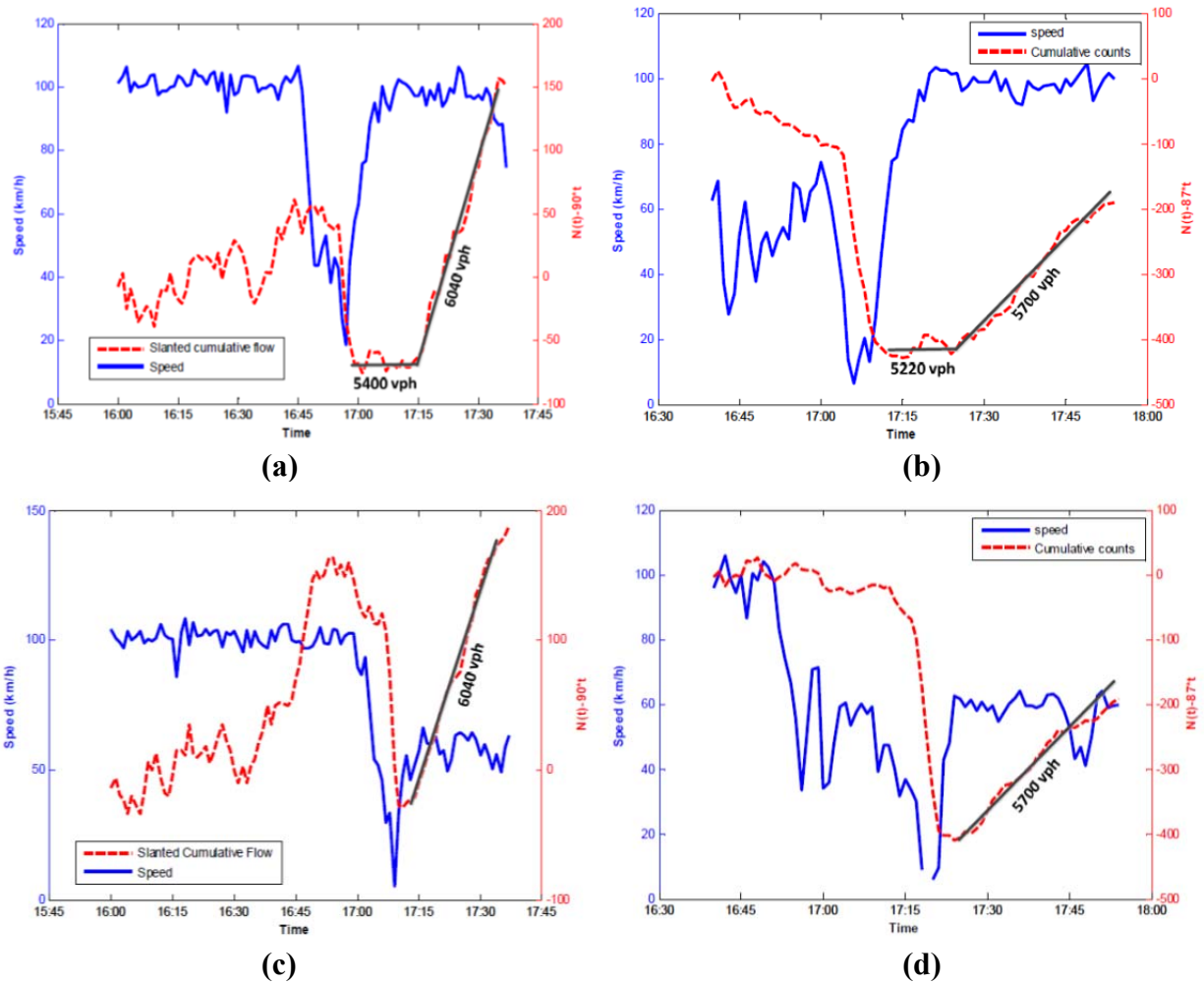


FIGURE 6 Average time mean speed (blue bold line) and slanted cumulative counts (red dash line) across three lanes at Location 1 and Location 8 on May 18, 2009 (a and c), and May 28, 2009 (b and d).

TABLE 1 Speed and Flow in Different Traffic State Points

	May 18, 2009		May 28, 2009	
	v_T (km/h)	q (veh/h)	v_T (km/h)	q (veh/h)
State 2	13.44	2,182.50	6.34	885.00
State 5	98.73	5,400.00	98.52	5,220.00
State 4	30.80	6,040.00	29.18	5,700.00
State 6	98.32	6,040.00	98.24	5,700.00

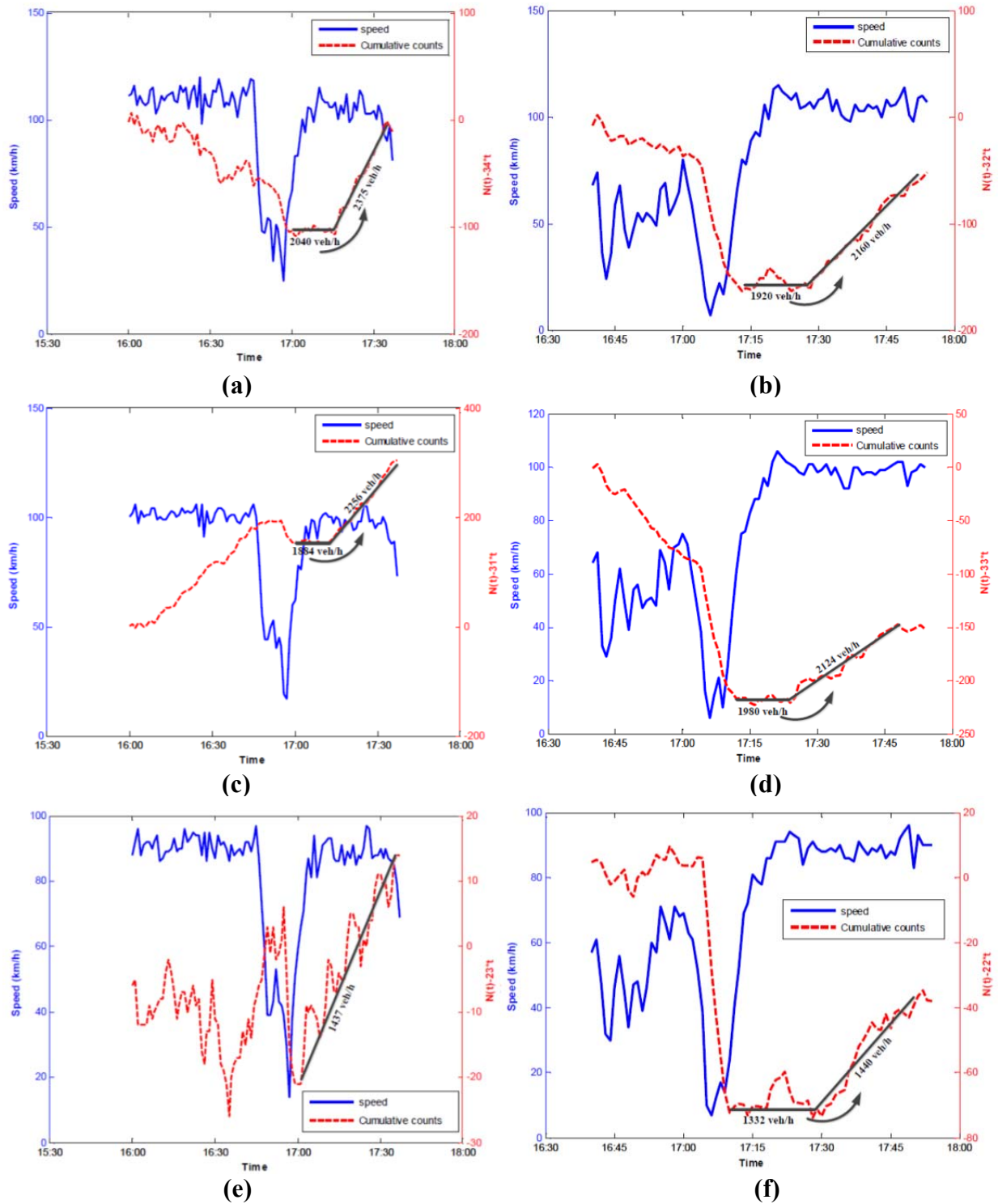


FIGURE 7 Speed and slanted cumulative count in each lane on May 18, 2009 (a, c, and e), and May 28, 2009 (b, d, and f), at Location 1. Flows are shown next to the coinciding slanted cumulative counts (bold black lines).

The 2 days also show a difference: the outflow of the standing queue fluctuates more with time on May 18 (see Figure 7e) whereas on May 28 (see Figure 7f) the outflow of the standing queue fluctuates strongly in the shoulder lane. At the moment is it unclear what could be the reason for the fluctuation.

When the bottleneck has been active, there are several different traffic states in the downstream of the bottleneck. Along the distance, the density decreases. Therefore, in the targeted scenario, a large range of density can be detected, which can reveal the flow distribution as a function of density across lanes. The flow distributions are shown in Figure 8. Red lines show the fast lane (median lane), black lines show the center lane, and blue lines show the slow lane (shoulder lane). Three bold lines (see Figure 8a and 8b) represent average flow distribution at three lanes based on all data. Circles and triangles are the empirical data collected in each lane at location 1 (see Table 1 and Figure 7). Those circles and triangles stand for the state of the outflow in each lane at Location 1 [i.e., State 5 and State 6, respectively (see Figure 1)]. The thin lines (in Figure 8c and 8d) represent the flow distributions at each location. The lines with five-point stars stand for the distribution at Location 8.

Figure 8 parts a and b show flow distributions on 2 different days. Both parts show a common feature. When the density lies within the range of 22 to 60 veh/km, the flow in the center lane is higher than that in both other lanes, although it keeps decreasing as density grows. When the density is around 60 veh/km, the fraction of the flow at shoulder lane reaches the minimum at around 23%. For the shoulder lane the decrease of the fraction of the flow was sharp, but afterwards the increase is only marginal. Meanwhile from 60 veh/km the fraction of the flow in median lane stops increasing with density and begins to stabilize at around 38%. Note that the density of 60 veh/km corresponds to a typical critical density.

When the density exceeds 132 veh/km (May 18) and 95 veh/km (May 28), the fraction of the flow in the median is almost equal to the fraction of the flow in the center lane, at around 35% for each, while the flow percentage at shoulder lane is around 30%. So even in states with a very high density, flow in the shoulder lane is still lower than that in the other lanes. When density reaches up to 220 veh/km, the flow begins to be distributed evenly over three lanes on May 18, while the flow distribution is more unstable on May 28. It is not surprising because in extremely high density situation standing vehicles can lead to some detection problems.

Figure 8 parts c and d show the flow distribution at 8 locations. The flow distribution in the median lane (red line) at Location 8 (marked as red five-point stars) is much higher than that at the other locations (see Figure 8c and 8d). In contrast, the flow distributions in the center and median lanes at Location 8 are the lowest. That is because vehicles merge into the median lane when passing through the lane-drop bottleneck. In the downstream of Location 8, the flow distribution in the median lane is lower than that at Location 8. For the other locations, the distribution situations are similar to each other. The authors explain this by the following: Vehicles force themselves into the traffic stream, and it takes some time and distance before equilibrium distribution sets in again. Therefore, it is believed that a high percentage of vehicles choose to leave the median lane by changing lanes between Location 8 and Location 7. This situation is only visible when the density reaches up to 130 veh/km.

The authors assume that because of the Keep Right Unless Overtaking rule in the Netherlands, the shoulder lane (slow lane) among three lanes is first choice for drivers when the density is extremely low. As the density increases to around 20 veh/km, the occupation of the center lane begins to be higher than that of the shoulder lane. The use of the median lane (fast

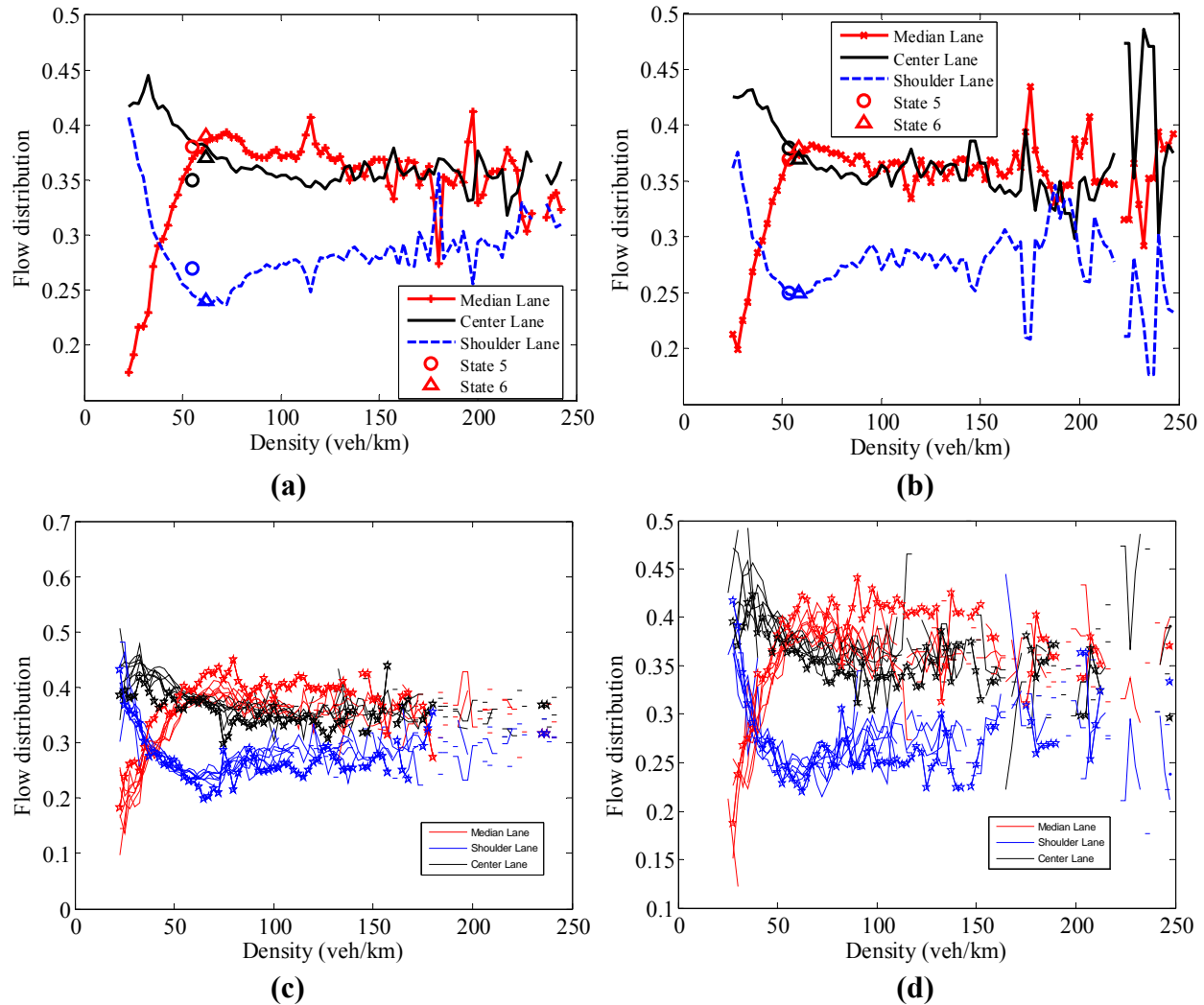


FIGURE 8 Flow distributions at different densities at three-lane freeway section. Parts *a* and *b* show average flow distributions over three lanes: median lane (red), center lane (black), and shoulder lane (blue) on 2 days, May 18 (*a* and *c*) and May 28 (*b* and *d*). Circles and triangles show the performance of each lane in State 5 and State 6, respectively, corresponding to data in Figure 7. Parts *c* and *d* show flow distributions at each of the eight locations. Each thin line shows a flow distribution at each location. Five-point stars represent the flow distribution at Location 8.

lane) is the least at that time. As the density increases, in contrast to the shoulder lane where flow fraction reduces considerably, the use of median lane grows sharply. Finally, the median lane and center lane are highly made use of while the shoulder lane is being underused.

Flow Distribution Over Lanes

Figure 9 shows the speed in each lane at the same average density over three lanes. As a result of the Keep Right Unless Overtaking rule, when the density is low the speed decreases from median lane to shoulder lane. The median lane is the fastest lane. As the density increases, the speed

becomes more equal among the lanes. When the flow distribution is equal for the median and center lanes, the speed is equal for the three lanes. Because in congestion the speeds are equal in all lanes, so the low flow in the shoulder lane must be because of a low density or large spacing. That means that microscopically in congestion the spacing between successive vehicles in the shoulder lane is the largest among three lanes.

Figure 10 shows the flow distributions in the four-lane freeway section upstream the lane-drop bottleneck. Note that the outflow of the upstream four-lane freeway section is the

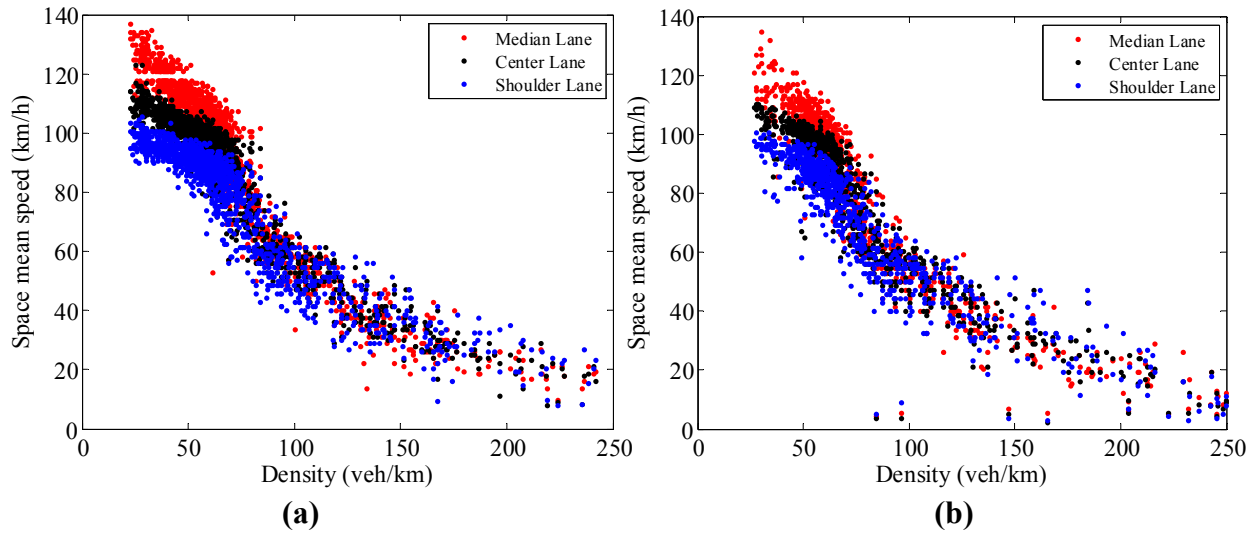


FIGURE 9 Speed–density plot in each lane in the three-lane section on 2 study days: (a) May 18 and (b) May 28. The density is the average density over three lanes.

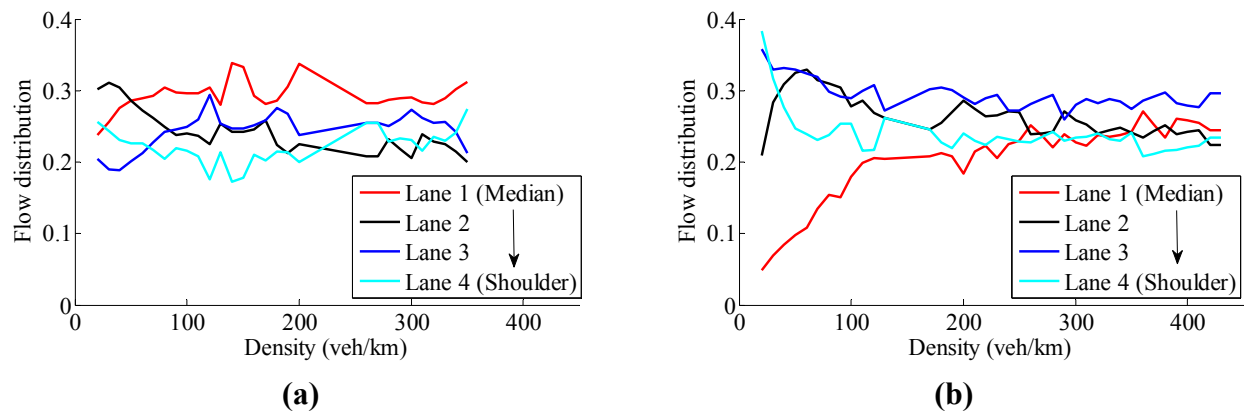


FIGURE 10 Flow distributions at different densities at four-lane freeway section on May 18 (a). The distribution on May 28 (b) is similar. The traffic flow is moving from Location 10 (a) to Location 9 (b).

inflow of the downstream three-lane freeway section. There are two locations for the data collection: Location 9 and Location 10 in Figure 3. Traffic flow moves from Location 10 to Location 9. The figure only shows the data for May 18; the data for May 28 are similar. In fact, the authors can distinguish two pairs of lanes. First, Lane 1 and 2 are the median and shoulder lane of one of the upstream branches of the road. The flow distributions in Lanes 3 and 4 are similar to that of Lanes 1 and 2, respectively, also originating from a two-lane road upstream. The flow distribution at two of the locations differs considerably. On one hand, in contrast to Location 10, which is in the upstream of the Location 9, Location 9 shows a lower flow in the median lane, especially for low densities. On the other hand, at Location 9 the flow in the shoulder lane is higher for low densities.

The noncompensated number of lane changes can be estimated by the difference in flow per lane between the two detectors for a certain density (e.g., one can see how much lower the flow is). Compensation is possible by other vehicles making opposite movements (e.g., vehicles moving into the lane). In Lane 3, the right center lane, the flow is higher at Location 9. Downstream of Location 9, all vehicles in the median lane have to merge into Lane 2. Drivers in Lane 2 (the left center lane) might anticipate this and make space for the drivers merging from the median lane. These lane changes can be considered as an explanation for the changes in lane flow distribution observed between Location 10 and Location 9. The relative flow in Lane 2 does not change as much, because there is a similar amount of lane changing from the median lane to Lane 2; what is observed is a decrease of the use of the median lane. The number of lane changing decreases as the average density over lanes increases. The flow distribution at Lane 2 and Lane 4 is nearly stable for both locations and study days. At Location 9 near the bottleneck, the flow in Lane 3 is always the highest for both study days. Note that the demand in the two upstream two-lane freeway sections could possibly greatly influence the flow distribution at Location 10.

CONCLUSIONS

This paper compares the downstream states of a stop-and-go wave with that of a standing queue. The standing queue in this paper is induced at a lane-drop bottleneck by a stop-and-go wave. Therefore, at one bottleneck there are two different congestion states observed. In the downstream of the congestion there are free-flow states, which means the two outflows detected downstream of congestion are the capacities of the road section. This paper applies shockwave analysis to find those two outflows at the same road section, which is well traceable in the real data. The most important finding is that the outflow of stop-and-go waves is much lower than that of a standing queue. Therefore, the capacity with congestion upstream can vary in a rather wide range (e.g., from 5,220 to 6,040 veh/h at a three-lane road section). The various capacities could be related to congestion states.

In the acceleration from stop-and-go waves, the detected flow grows as the speed increases. In contrast, in the acceleration from standing queues, the flow remains a constant flow along freeway. So during the acceleration from a standing queue, the density there is an inverse proportional function of speed.

There are two other findings. First, different features of outflow from congestion in different lanes can be found. Strong fluctuations occasionally can be observed in the shoulder lane, which might even trigger stop-and-go waves later on, for instance near a next bottleneck. Second, the flow distribution over three lanes is presented. This shows that particularly near the head of a

standing queue more vehicles can merge into the lane adjacent to the ending lane, thereby locally increasing the capacity of that lane. The capacity of the shoulder lane is markedly wasted when in congestion. The reason for the low flow distribution in the shoulder lane is the large spacing between successive vehicles.

Future research should show the mechanisms behind these features from a behavioral perspective (whether people behave differently), from a vehicle perspective (what the influences of different acceleration profiles are), or from a flow perspective (what for instance the influence of voids is).

ACKNOWLEDGMENTS

This research is financially supported by the China Scholarship Council, the NWO grant for “There is plenty of room in the other lane,” and the visiting scholar grant from the TU Delft Transport Institute of Delft University of Technology.

REFERENCES

1. Hall, F. L., and K. Agyemang-Duah. Freeway Capacity Drop and the Definition of Capacity. In *Transportation Research Record 1320*, TRB, National Research Council, Washington, D.C., 1991, pp. 91–98.
2. Cassidy, M. J., and R. L. Bertini. Some Traffic Features at Freeway Bottlenecks. *Transportation Research Part B: Methodological*, Vol. 33, No. 1, 1999, pp. 25–42.
3. Srivastava, A., and N. Geroliminis. Empirical Observations of Capacity Drop in Freeway Merges with Ramp Control and Integration in a First-Order Model. *Transportation Research Part C: Emerging Technologies*, Vol. 30, No. 0, 2013, pp. 161–177.
4. Chung, K., J. Rudjanakanoknad, and M. J. Cassidy. Relation Between Traffic Density and Capacity Drop at Three Freeway Bottlenecks. *Transportation Research Part B: Methodological*, Vol. 241, No. 1, 2007, pp. 82–95.
5. Cassidy, M. J., and J. Rudjanakanoknad. Increasing the Capacity of an Isolated Merge by Metering Its On-ramp. *Transportation Research Part B: Methodological*, Vol. 39, No. 1), 2005, pp. 896–913.
6. Oh, S., and H. Yeo. Estimation of Capacity Drop in Highway Merging Sections. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2286, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 111–121.
7. Kerner, B. S. Empirical Macroscopic Features of Spatial-Temporal Traffic Patterns at Highway Bottlenecks. *Physical Review E*, Vol. 65, No. 4, 2002, p. 046138.
8. Laval, J. A., and C. F. Daganzo. Lane-changing in Traffic Streams. *Transportation Research Part B: Methodological*, Vol. 40, No. 3, 2006, pp. 251–264.
9. Daganzo, C. F. A Behavioral Theory of Multi-lane Traffic Flow. Part I: Long Homogeneous Freeway Sections. *Transportation Research Part B: Methodological*, Vol. 36, No. 2, 2002, pp. 131–158.
10. Knoop, V. L., S. P. Hoogendoorn, and H. van Zuylen. Empirical Differences Between Time Mean Speed and Space Mean Speed. In *Traffic and Granular Flow '07* (C. Appert-Rolland et al., eds.), Springer Berlin Heidelberg, 2009, pp. 351–356.
11. Yuan, Y., J. W. C. van Lint, T. Schreiter, S. P. Hoogendoorn, and J. L. M. Vrancken. Automatic Speed-Bias Correction with Flow-Density Relationships. Presented at 2010 International Conference on Networking, Sensing and Control (ICNSC), 2010.
12. Ou, Q. Fusing Heterogeneous Traffic Data: Parsimonious Approaches using Data-Data Consistency. In *TRAIL Research School*, Delft University of Technology, the Netherlands, 2011.

FUNDAMENTAL DIAGRAM

An Overview of “Revisiting the Empirical Fundamental Relationship”

BENJAMIN COIFMAN
The Ohio State University

This paper presents an extended overview of a new methodology for deriving an empirical fundamental relationship from vehicle detector data. The new methodology seeks to address several sources of noise present in conventional measures of the traffic state that arise from the data aggregation process (e.g., averaging across all vehicles over a fixed time period). In the new methodology vehicles are no longer taken successively in the order in which they arrived, and there is no requirement to seek out stationary traffic conditions; rather, the traffic state is measured over the headway for each individual vehicle passage, and the vehicles are grouped by similar lengths and speeds before aggregation. Care is also taken to exclude measurements that might be corrupted by detector errors. The result is a homogeneous set of vehicles and speeds in each bin.

While conventional fixed time averages may have fewer than 10 vehicles in a sample, the new binning process ensures a large number of vehicles in each bin before aggregation. Researchers calculate the median flow and median occupancy for each combined length and speed bin. Then they connect these median points across all of the speed bins for a given vehicle length to derive the empirical fundamental relationship for that length. This use of the median is also important; unlike conventional aggregation techniques that find the average, the median is far less sensitive to outliers arising from uncommon driver behavior or occasional detector errors.

INTRODUCTION

This paper presents an extended overview of Coifman (1), which develops a new methodology for deriving an empirical fundamental relationship (FR) from vehicle detector data. This work is important because much of traffic flow theory depends on the existence of an FR between flow (q), density (k), and space–mean speed (v) either explicitly [e.g., Lighthill and Whitham (2), and Richards (3) hydrodynamic model] or implicitly [e.g., car following models (4, 5)]. The FR is commonly characterized in terms of a bivariate relationship between two of the three parameters (in each case the third parameter can be calculated from the fundamental equation, repeated in Equation 1). All of the empirically generated FRs use data that average conditions over time or space, or both, to calculate the traffic state (q, k, v). It is difficult to measure k directly, so many empirical FR studies use occupancy, occ , as a proxy for k , where occ is the percentage of the sampling period, T , that vehicles occupy the detector. As shown in Equation 2, in stationary traffic occ is proportional to k by the average effective vehicle length, L_{eff} during T ; where a given vehicle’s effective length is the sum of its physical length and the size of the detection zone.

$$q = k * v \quad (1)$$

$$occ = k * L_{eff} \quad (2)$$

Most empirical FR studies use traffic state measurements from conventional detectors that average vehicle measurements over fixed time sampling periods. Figure 1a shows the measured q versus occ from one of the dual loop detectors in the Berkeley Highway Laboratory (BHL) (6) on a single lane on a single day, using $T = 30$ -s aggregation periods. The results are typical of 30-s aggregation periods aside from the fact that this location experienced over 12 h of recurring congestion on this day, as per the corresponding time series speed shown in Figure 1d. The q versus occ plot shows considerable scatter, and it is hard to imagine any single curve that would be representative of all of the observed data points. The scatter is commonly attributed to combining nonstationary traffic states [e.g., Cassidy (7)], and debate continues as how best to address the scatter.

The choice of T is an effort to balance between maximizing the number of vehicles in the sample and minimizing the averaging across inhomogeneous traffic states. Often setting $T = 30$ s is considered to be a good balance between the two competing objectives, though the original use of $T = 30$ s appears to have been for the convenience of telecommunications (8). Figure 1b repeats the q versus occ for the same vehicles from Figure 1a, only now using $T = 5$ min. With the longer sample period, each data point in Figure 1b is more likely to combine different nonstationary traffic states, yet the scatter in this plot is diminished compared to $T = 30$ s.

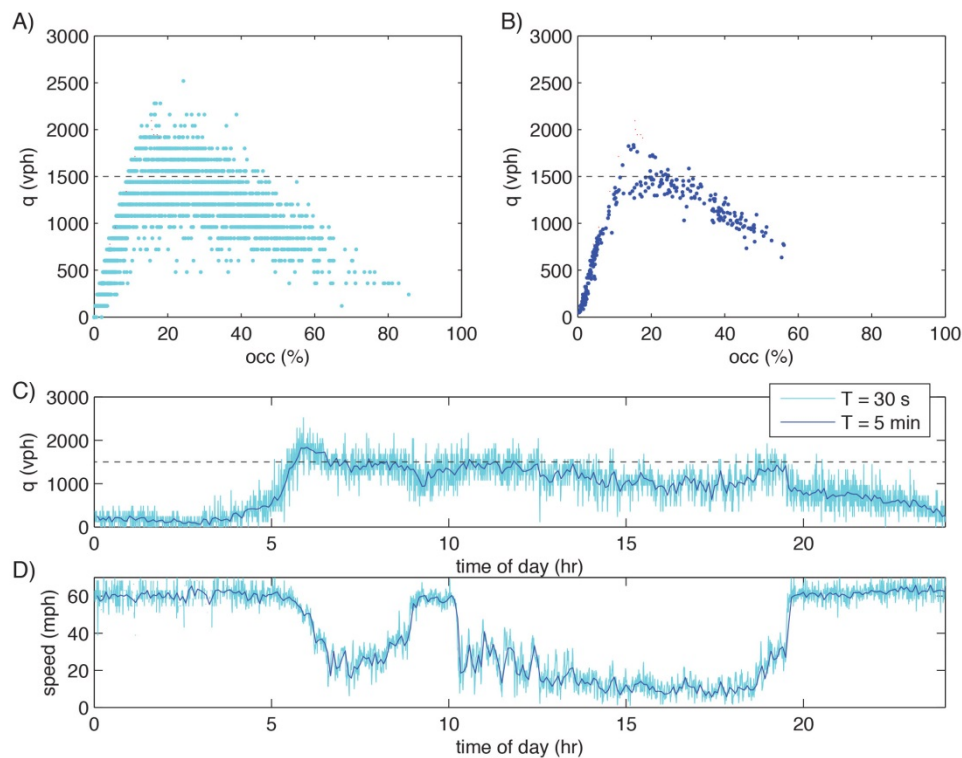


FIGURE 1 Conventional q versus occ : (a) $T = 30$ s, (b) $T = 5$ min. The corresponding time series: (c) flow and (d) speed. For reference, a dashed line is shown at $q_p = 1,500$ veh/h.

Thus, the noise in the $T = 30$ -s plot cannot strictly be due to averaging across nonstationary traffic states. The $T = 5$ -min plot also exhibits plateauing predicted by Hurdle and Datta (9) and subsequently illustrated in Hsu and Banks (10); whereby the q versus occ within the queue is truncated at q_p , some maximum q below the bottleneck capacity, due to traffic entering from on-ramps between the detector location and the bottleneck. This downstream demand consumes a portion of the bottleneck capacity, and in this case the resulting plateau in the detector data plot falls around $q_p = 1,500$ veh/h, as shown with a dashed line throughout Figure 1.

Conventional fixed time sampling is a crude methodology that was originally designed to smooth out variability across vehicles in an era when computing power was expensive. While it is true that successive vehicles will usually experience similar traffic conditions while traversing the detector, the impacts of differing vehicle properties can far outweigh any benefits that might come from arbitrarily grouping vehicles based on successive passages; for example, Coifman (11) and Coifman et al. (12) show empirically that the range of feasible vehicle length undermines conventional relationships between q and occ . Coifman (13) uses hypothetical microscopic models to revisit the process of generating empirical FR and uncovered several commonly underappreciated factors that result in surprisingly large, nonlinear distortions of empirical traffic state measurements.

Briefly reviewing Coifman (13), in general for a fleet of homogeneous vehicles, stationary traffic, in a sample with a large number of vehicles occ is related to k via Equation 2. If one assumes a triangular flow-density FR (denoted $qkFR$), the curve is uniquely defined by capacity, q_o , free speed, v_f , and jam density, k_j . Then, extending to the flow-occupancy FR ($qoccFR$), Equation 3 gives jam occupancy, occ_j . The resulting $qoccFR$ is shown on the right-hand side of Figure 2a for a hypothetical example with $v_f = 65$ mph, $q_o = 2,400$ veh/h, $k_j = 211$ veh/min, and $L_{eff} = 20$ ft, while the left-hand side shows the corresponding speed-flow relationship ($qvFR$), transposed from the commonly used orientation to facilitate direct comparisons between these two forms of the FR.

$$occ_j = k_j * L_{eff} \quad (3)$$

Even under strictly stationary traffic conditions with homogeneous vehicles, Coifman (13) found that conventional aggregated q , occ , and v measurements should exhibit large scatter in the queued regime arising from a combination of

1. Errors due to a non-integer number of vehicle headways in a given sample,
2. Averaging over a small number of vehicles during low q ,
3. The inclusion of detector errors, and
4. The mixing of inhomogeneous vehicles within a sample.

Generally the errors grow larger at lower v . Coifman (13) also found that, unfortunately, the q - occ plane is skewed such that the noisy low-speed samples cover a disproportionately large area. The points in the $qvFR$ on the left-hand side of Figure 2a are plotted at 5-mph intervals, and the corresponding states are shown in the $qoccFR$ by projecting horizontally to the curve on the right-hand side of the figure. Proportionately the $qoccFR$ greatly distorts the relationship to v . The dashed lines in Figure 2a show that when speed has dropped from v_f by 38% ($v = 40$ mph) flow has only dropped from q_o by 10%, and when speed has dropped by 85% ($v = 10$ mph) flow has only dropped by 49%. In other words, the higher-speed data are compressed into a narrow

sliver; for example, one can see that the highest speeds in the queued regime are compressed to a small range of q (or of occ), while the low-speed data ($v = 10$ mph) are spread over more than half of the feasible range of q or occ .

METHODOLOGY

Coifman (1) develops a new sampling methodology designed to address the above shortcomings by specifically focusing on the sources of the distortions and minimizing their impacts. In this approach vehicles are no longer taken successively, rather, they are sorted as follows:

- To eliminate the impacts of noninteger headways, individual vehicle headway, h , is measured instead of q . The associated detector on-time is measured and is used in conjunction with h instead of occ . Here h is measured rear bumper to rear bumper to ensure the driver's chosen gap is combined with his or her vehicle's on-time.
 - Speed and vehicle length are measured for each individual vehicle.
 - To minimize the impacts arising from a large range of vehicle lengths, the vehicles are sorted into length bins that only span 5 or 10 ft.
 - To minimize the impacts of different speeds, the vehicles are then sorted into speed bins that only span 1 mph.
 - To minimize the impacts of detector errors, any vehicle following an unmatched pulse, involved in a suspected pulse breakup, or following a suspected pulse breakup is excluded, because its measured h may be inaccurate.
 - Finally, to ensure the largest possible number of vehicles per sample, the median headway and median on-time are found for each bin from the subsection of Step 3 by Step 4 above, and converted to q and occ via Equations 4 and 5. This clustering approach groups vehicles based on similar speed and length.

$$q = \frac{1}{\text{median}(h)} \quad (4)$$

$$occ = \frac{\text{median}(on)}{\text{median}(h)} \quad (5)$$

ANALYSIS AND DISCUSSION

The clean hypothetical FR curves in Figure 2a were derived from a homogeneous fleet of vehicles with $L_{\text{eff}} = 20$ ft under stationary conditions, over large sample periods. If the homogeneous fleet has a different L_{eff} , it yields a different curve. To extend the model from Coifman (13) to homogeneous vehicles with longer L_{eff} , one must recognize that q_o and k_j are functions of L_{eff} . If one assumed that all vehicles came to a stop with a constant physical gap between vehicles, the parameters of $qkFR$ scale as a function of L_{eff} . Using several different values of L_{eff} to recalculate the FR gives rise to a family of curves as shown in Figure 2b. The greater L_{eff} , the lower the curve is in this plot, with the top curve corresponding to $L_{\text{eff}} = 20$ ft and the bottom to $L_{\text{eff}} = 73$ ft.

Coifman (1) applies the process outlined above in the Methodology section to 18 successive days of empirical detector data from one directional detector station in the BHL with four lanes. Combining all of the individual vehicle actuations from all four lanes over 18 days at the detector station, measured length is used to sort each vehicle into one of 10 different length bins and individual vehicle speed to further sort these vehicles into one of 70 speed bins. After sorting all vehicles in this manner, Equations 3 and 4 are applied for each one of these combined length and speed bins. Finally, the $qoccFR$ is found by connecting the q and occ values across successive speed bins for a given length bin. The right-hand side of Figure 2c shows the resulting curves by length bin subject to the following three exclusions:

1. To remove the impacts of the downstream inflow, all speeds in all length bins that result in q above the threshold q_p for the 18- to 22-ft length bin are suppressed (as per the corresponding $qvFR$ in the left-hand side of Figure 2c).
2. To remove the impacts of the measurement errors at low speeds, all speeds below 5 mph in all length bins are suppressed.
3. To remove the impacts of small sample sizes, all combined length and speed bins with fewer than 100 vehicles are suppressed.

As a result of the third exclusion, only seven length bins remain with sufficient data. The seven curves are distinctly visible in the $q - v$ plane on the left-hand side of Figure 2c. The length range increases from the top curve to the bottom curve in this plot. On the right-hand side the two shortest length bins remain distinct, but the curves from the five longest bins overlap, in part due to the vertical compression as L_{eff} increases, evident in Figure 2b; in part due to smaller sample sizes in these length bins; in part due to the fact that the speed measurement errors from Exclusion 2 are greater for longer vehicles simply because they are over the detector for a longer amount of time; and in part due to the fact that the actual speed-occupancy relationship appears to exhibit a slight dependency on L_{eff} for these longer vehicles, leading to a small lateral shift in the right-hand side of Figure 2c. All of the curves exhibit trends consistent with the hypothetical example in Figure 2b. Although the speed range on Figure 2c is relatively small, 5 to 17 mph, the occ range spans roughly a quarter of the observable values below jam occupancy.

The dotted curves in Figure 2d relax Exclusion 1, lifting the upper bound speed exclusion to 50 mph and eliminates the lower bound speed, Exclusion 2, altogether (for reference, the curves from Figure 2c are repeated in bold). In the right-hand side of Figure 2d the measured $qoccFR$ curves for most of the length bins flatten out at higher q (corresponding to higher v), in other words, in the region above the threshold q_p from the downstream inflow, shown with a dashed line in these plots. For the shortest length bin (again, accounting for over two-thirds of the vehicles) the curve now extends measurements down to 1 mph. The curve remains roughly straight throughout this low-speed region, possibly indicating that with a sufficiently large sample size the median is not very sensitive to the outliers arising from the dual loop detector measurement limitations for these very low speeds. If so, the resulting occ range for speeds between 1 and 4 mph covers an additional quarter of the feasible values of occ , with the combined speed range from 1 mph to 17 mph covering more than half of the feasible occ values.

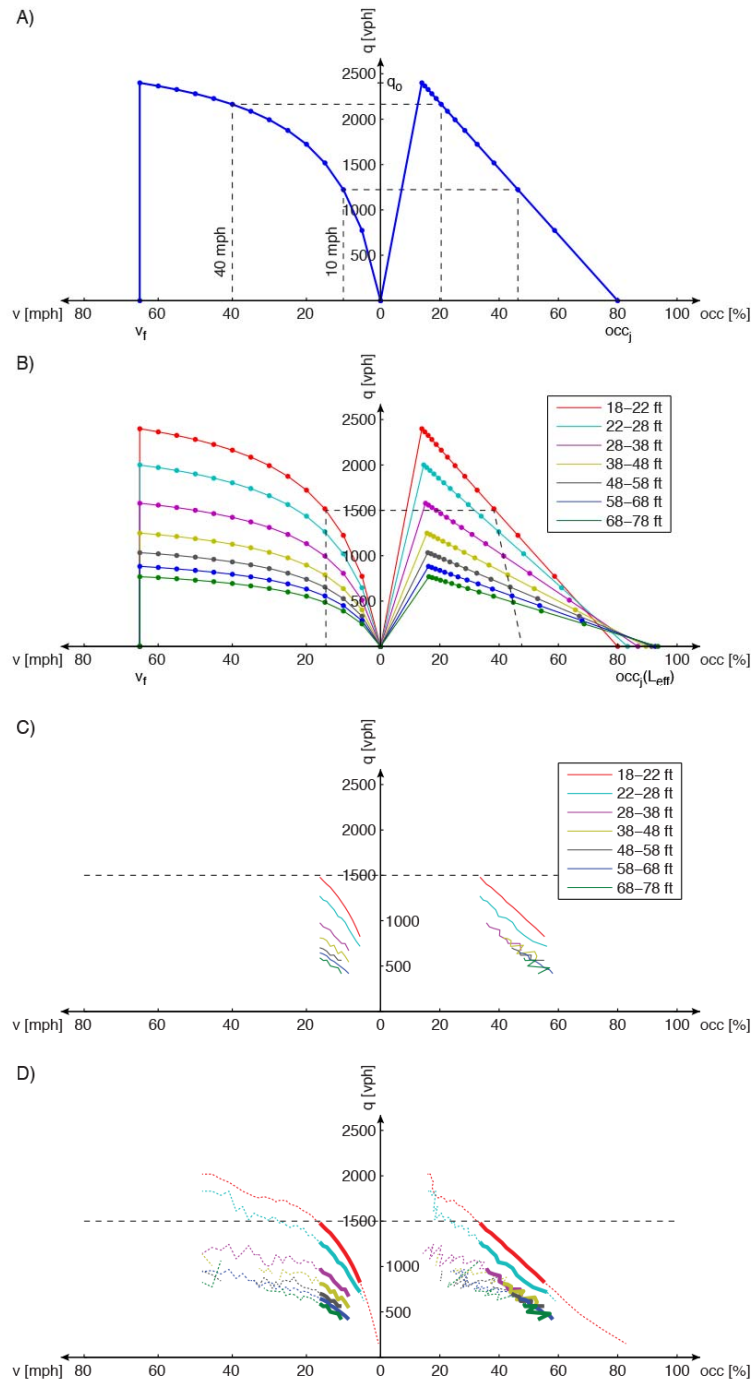


FIGURE 2 (a) Hypothetical $qvFR$ (left) and $qoccFR$ (right) for a homogeneous fleet of passenger vehicles with $L_{eff} = 20$ ft under stationary conditions and very long sampling periods and (b) extending the relationship to homogeneous fleets with longer lengths. The empirical $qvFR$ and corresponding $qoccFR$ for each length bin individually using the data from all 18 days and all four lanes combined together for (c) 5 mph to 17 mph and (d) 1 to 50 mph. A dashed line is shown at $q_p = 1,500$ veh/h for reference.

The distortions of the q -occ plane discussed in the context of Figure 2a are readily apparent in Figure 2 parts *c* and *d* as follows. The speed range in Figure 2c spans roughly 18% of the feasible speeds ($5 \text{ mph} \leq v \leq 17 \text{ mph}$), but these data span roughly 25% of the feasible q and occ. As shown in Figure 2d, if one extends the range to $1 \text{ mph} \leq v \leq 17 \text{ mph}$ (roughly 26% of the feasible speeds) the data now span more than 50% of the feasible q and occ measurements.

CLOSING

This paper provided an extended overview of Coifman (*1*). Greater details of the process and discussion of the implications can be found in that paper.

REFERENCES

1. Coifman, B. Revisiting the Empirical Fundamental Relationship, *Transportation Research Part B*, Vol. 68, 2014, pp. 173–184.
2. Lighthill, M., and G. Whitham. On Kinematic Waves II. A Theory of Traffic Flow on Long Crowded Roads. *Proc. Royal Society of London, Part A*, Vol. 229, No. 1178, 1955, pp. 317–345.
3. Richards, P. Shock Waves on the Highway. *Operations Research*, Vol. 4, No. 1, 1956, pp. 42–51.
4. Chandler, R. E., R. Herman, and E. W. Montroll. Traffic Dynamics: Studies in Car Following. *Operation Research*, Vol. 6, No. 2, 1958, pp. 165–184.
5. Gazis, D. C., R. Herman, and R. W. Rothery. Nonlinear Follow the Leader Models of Traffic Flow. *Operations Research*, Vol. 9, No. 4, 1961, pp. 545–567.
6. Coifman, B., D. Lyddy, and A. Sabardonis. The Berkeley Highway Laboratory-Building on the I-880 Field Experiment. *Proc., IEEE ITS Council Annual Meeting*, IEEE, 2000, pp. 5–10.
7. Cassidy, M. Bivariate Relations in Nearly Stationary Highway Traffic. *Transportation Research Part B*, Vol. 32, No. 1, 1998, pp. 49–59.
8. Gazis, D., and R. Foote. Surveillance and Control of Tunnel Traffic by an On-Line Digital Computer. *Transportation Science*, Vol. 3, No. 3, 1969, pp. 256–275.
9. Hurdle, V. F., and P. K. Datta. Speeds and Flows on an Urban Freeway: Some Measurements and a Hypothesis. In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp. 127–137.
10. Hsu, P., and J. Banks. Effects of Location on Congested-Regime Flow-Concentration Relationships for Freeways, In *Transportation Research Record 1398*, TRB, National Research Council, Washington, D.C., 1993, pp. 17–23.
11. Coifman, B. Improved Velocity Estimation Using Single Loop Detectors. *Transportation Research Part A*, Vol. 35, No. 10, 2001, pp. 863–880.
12. Coifman, B., S. Dhoorjaty, and Z. Lee. Estimating Median Velocity Instead of Mean Velocity at Single Loop Detectors. *Transportation Research Part C*, Vol. 11, No. 3-4, 2003, pp. 211–222.
13. Coifman, B. Jam Occupancy and Other Lingering Problems with Empirical Fundamental Relationships. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2422, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 104–112.

FUNDAMENTAL DIAGRAM

Universalities in Fundamental Diagrams of Cars, Bicycles, and Pedestrians

A. SEYFRIED

M. BOLTES

ST. HOLL

W. MEHNER

J. ZHANG

Jülich Supercomputing Centre, Germany

E. ANDRESEN

Bergische Universität, Germany

A. SCHADSCHNEIDER

Universität zu Köln, Germany

Since the pioneering work of Greenshields the fundamental diagram is used to characterize and describe the performance of traffic systems (1, 2). During the last years the discussion and growing database revealed the influence of human factors, traffic types, or ways of measurements on this relation (3, 4). The manifoldness of influences is important and relevant for applications but moves the discussion away from the main feature characterizing the transport properties of traffic systems. The authors focus again on the main feature by comparing the fundamental diagram of cars, bicycles, and pedestrians moving in a row in a course with periodic boundaries. The underlying data are collected by three experiments, performed under well-controlled laboratory conditions (5–9). In all experiments the setup, in combination with technical equipment or methods of computer vision, allowed the authors to determine the trajectories with high precision. The trajectories visualized by space-time diagrams show three different states of motion (free-flow state, jammed state, and stop-and-go waves) in all these systems. Obviously the values of speed, density, and flow of these three systems cover different ranges. However, after a simple rescaling of the velocity by the free speed and of the density by the length of the agents, the fundamental diagrams conform regarding the position and height of the capacity. This indicates that the similarities between the systems go deeper than expected and offers the possibility of a universal model for heterogeneous traffic systems.

EXPERIMENTS

All three experiments were performed with similar setups, namely on circuits with closed boundary conditions where only single-file motion was allowed. Series of runs were carried out with a maximal number of participants $N_{\max} = 70, 23,$ and 33 for the pedestrian, car, and bicycle experiment, respectively. To adjust, the global density different runs were performed with different numbers of participants N . In general, participants were asked to move normally without overtaking. Details of the experiments could be found in Sugiyama et al. (5) and Tadaki et al. (6)

for cars, in Andresen et al. (7) and Zhang et al. (8) for bicycles, and in Seyfried et al. (9, 10) for pedestrian. Time-space diagrams are shown in the center of Figures 1 and 2. Similar plots for cars can be found in Sugiyama et al. (5). In all three cases a transition from free flow to jammed flow can be observed when the global density is increased. In the free-flow regime all agents can move with their desired speed, whereas in the jammed regime stop-and-go waves are observed.

On the right of Figures 1 and 2 the density–flow relations are shown. Details of the measurement method could be found in Zhang et al. (8). The fundamental diagram of pedestrians shows three regimes $\rho \in [0, 1.0] \text{ m}^{-1}$, $[1.0, 1.7] \text{ m}^{-1}$, and $[1.7, 3.0] \text{ m}^{-1}$ corresponding to three states of pedestrian movement. At the free-flow regime ($\rho < 1.0 \text{ m}^{-1}$) the flow increases monotonically with the density. For the congested state ($\rho > 1.0 \text{ m}^{-1}$) the specific flow starts to decrease with increasing density. For $\rho > 1.7 \text{ m}^{-1}$ stop-and-go waves dominate the motion of the pedestrians (9, 10). Similar results are observed in the bicycle system (7, 8).

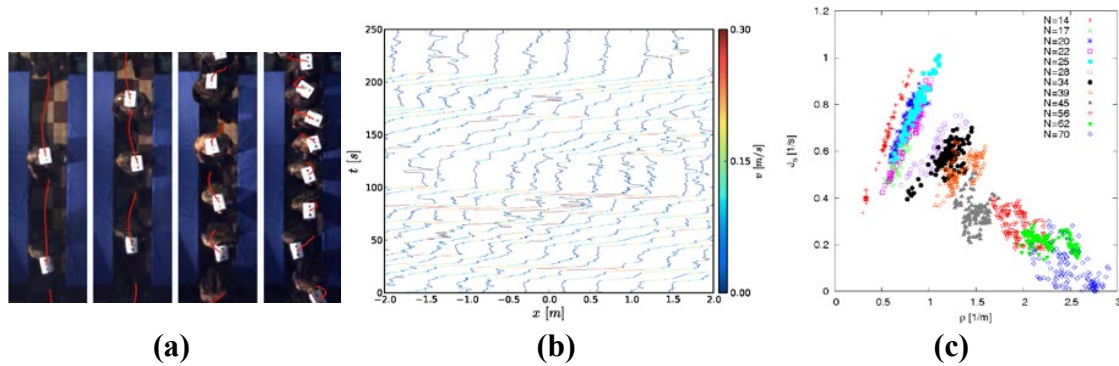


FIGURE 1 (a) Snapshots of the pedestrian experiment at different densities. The red lines show the trajectories determined automatically from video recordings; (b) trajectories in space and time in the measurement area (of length 4 m) with $N = 70$; and (c) density–flow relation. Colors indicate data for runs with different N .

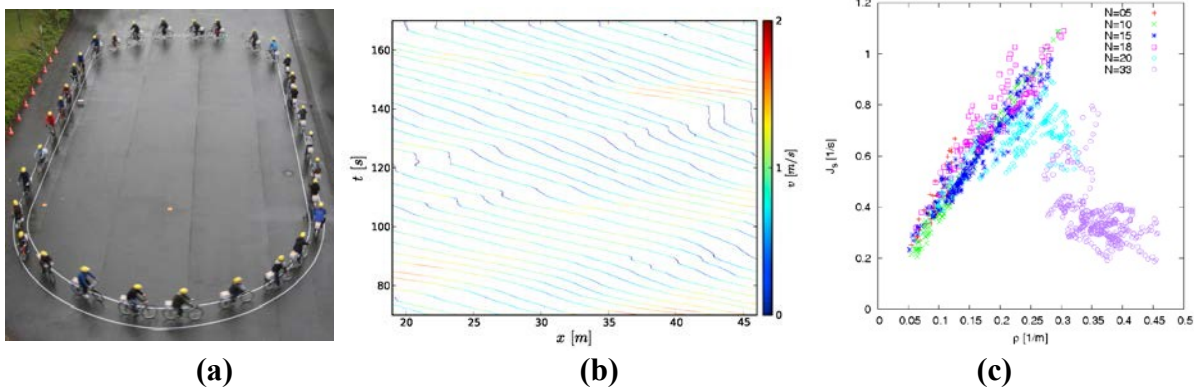


FIGURE 2 (a) Snapshots of the bicycle experiment; (b) trajectories in the measurement area (of length 27 m) for the bicycle experiment with $N = 33$. The same structures can be found in trajectories of vehicle systems (5, 6). (c) Density–flow relation. Colors indicate data for runs with different N .

RESULTS

Plotting the fundamental diagram of these three systems in one diagram shows that the data points occupy different ranges of density as well as speeds and do not seem to be comparable to each other (see Figure 3). To take into account the different scales of sizes and speeds of the agents the authors rescale these quantities. For the length of the agents the authors use $L_0(p) = 0.4$ m for pedestrians, $L_0(c) = 3.9$ m for cars (6), and the mean value of $L_0(b) = 1.73$ m for bicycles (7, 8). For scaling the speed the authors used the free-flow speed of each agent. From measurements of the free-flow speed in special runs of the experiments it is known that they are about 1.4 m/s for pedestrians and 5.5 m/s for bicycles. For cars the authors use 11.1 m/s (about 40 km/h) according to Tadaki et al. (6).

After rescaling it is found that the fundamental diagrams agree, and in all three cases the free-flow regimes end at approximately $\rho^*L_0 = 0.5$ (see Figure 4). This implies that the transition to the congested state occurs when nearly 50% of the available space is occupied. Moreover, the capacity (i.e., the maximal flow) agrees for the three systems after the rescaling and amounts 0.25 to 0.30. In the congested regime the slopes of the fundamental diagram are again similar for all three systems.

The transport properties in such systems could be approximated by the universal equation $\tilde{v} = 1 - \tilde{\rho}$ with $\tilde{v} = v/v_0$ and $\tilde{\rho} = \rho L_0$. The normalized maximal flow is then 0.25 at a relative density of 0.5. This corresponds to the properties to the ASEP (11, 12), which for a long time has been considered a minimal model for traffic flows. The main feature of this model is volume exclusion. Also models for pedestrian dynamics (13–15) show that these transport characteristics could be reproduced by an appropriate consideration of a velocity-dependent volume exclusion, which seems to be a universal characteristic of such systems. Considering this universality the authors conclude that other properties of the agent, like acceleration or inertia, are less relevant for the structure of the fundamental diagram in single-file traffic systems of different agent types. In other words models without a proper consideration of the volume exclusion miss an important aspect of traffic systems.

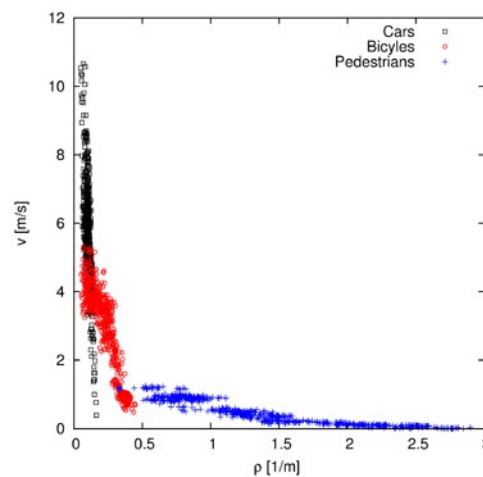


FIGURE 3 Fundamental diagrams for cars, bicycles, and pedestrians. Raw data of the density–speed relation.

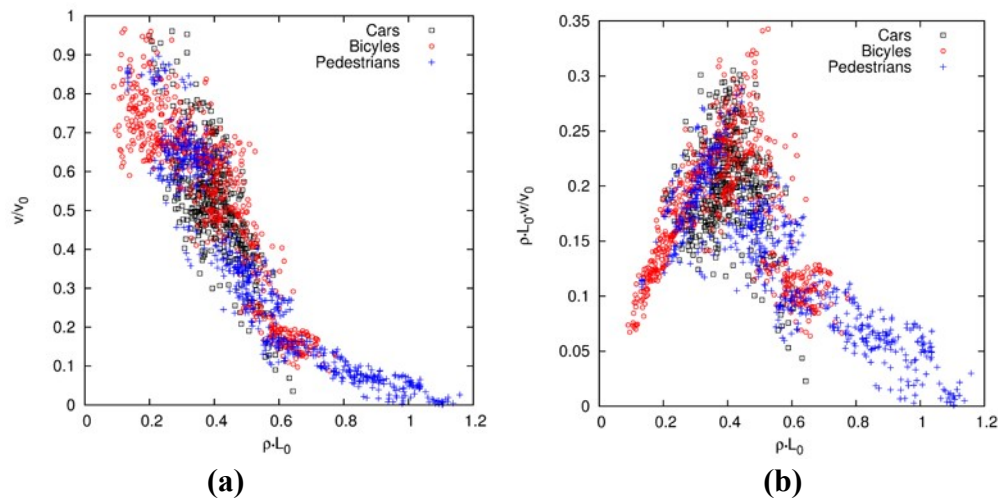


FIGURE 4 Fundamental diagrams for cars, bicycles, and pedestrians: (a) scaled data of the density–speed relation; speed and density rescaled with the free speed v_0 and the length of the agents L_0 ; (b) scaled density–flow relation; after the scaling the fundamental diagrams agree in the density range observed.

REFERENCES

1. Greenshields, B. D. The Photographic Method of Studying Traffic Behavior. *Highway Research Board Proceedings*, Vol. 13, 1933.
2. Greenshields, B. D. A Study of Highway Capacity. *Highway Research Board Proceedings*, Vol. 14, 1935, pp. 448–477.
3. Coifmann, B. Jam Occupancy and Other Lingering Problems with Empirical Fundamental Relationships. Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.
4. Hranac, R., E. Sterzin, D. Krechmer, H. Rakha, and M. Farzaneh. *Empirical Studies on Traffic Flow in Inclement Weather*. Report Number: FHWA-HOP-07-073. FHWA, U.S. Department of Transportation, 2006.
5. Sugiyama, Y., M. Fukui, M. Kikuchi, K. Hasebe, A. Nakayama, K. Nishinari, S. Tadaki, and S. Yukawa. Traffic Jams Without Bottlenecks: Experimental Evidence for the Physical Mechanism of the Formation of a Jam. *New Journal of Physics*, Vol. 10, 2008, p. 033001.
6. Tadaki, S., M. Kikuchi, M. Fukui, A. Nakayama, K. Nishinari, A. Shibata, Y. Sugiyama, T. Yosida, and S. Yukawa. Phase Transition in Traffic Jam Experiment on a Circuit. *New Journal of Physics*, Vol. 15, 2013, p. 103034.
7. Andresen, E., M. Chraïbi, A. Seyfried, and F. Huber, Basic Driving Dynamics of Cyclists. Proc., Simulation of Urban Mobility 2013, Behrisch M., M. Knocke, and D. Krajzewicz (eds.), *Lecture Notes in Computer Science (LNCS)*, Vol. 8594, Springer, 2014, pp. 18–32.
8. Zhang, J., W. Mehner, S. Holl, M. Boltes, E. Andresen, A. Schadschneider, and A. Seyfried. Universal Flow-Density Relation of Single-file Bicycle, Pedestrian, and Car Motion. Submitted to *New Journal of Physics*, 2014.
9. Seyfried, A., M. Boltes, J. Kähler, W. Klingsch, A. Portz, T. Rupperecht, A. Schad-Schneider, B. Steffen, and A. Winkens. Enhanced Empirical Data for the Fundamental Diagram and the Flow Through Bottlenecks. In *Pedestrian and Evacuation Dynamics 2008*, Springer, 2010, pp. 145–156.
10. Seyfried, A., A. Portz, and A. Schadschneider. Phase Coexistence in Congested States of Pedestrian

- Dynamics. Bandini, S., S. Manzoni, H. Umeo, and G. Vizzari (Eds.), *Cellular Automata*, Springer, Vol. 6350, 2010, pp. 496–505.
11. Derrida, B. An Exactly Soluble Non-equilibrium System: The Asymmetric Simple Exclusion Process, *Phys. Rep.* Vol. 301, 1998, pp. 65–83.
 12. Schadschneider, A., D. Chowdhury, and K. Nishinari. *Stochastic Transport in Complex Systems: From Molecules to Vehicles*. Elsevier, 2010.
 13. Seyfried, A., B. Steffen, and T. Lipper. Basic of Modeling the Pedestrian Flow. *Physica A*, Vol. 368, 2006, pp. 232–238.
 14. Chraïbi, M., A. Seyfried, and A. Schadschneider. Generalized Centrifugal Force Model for Pedestrian Dynamics. *Physical Review E*, Vol. 82, 2010, p. 046111.
 15. Kirchner, A., H. Klüpfel, K. Nishinari, A. Schadschneider, and M. Schreckenberg. Discretization Effects and the Influence of Walking Speed in Cellular Automata Models for Pedestrian Dynamics. *Journal of Statistical Mechanics*, No. 10, 2004, P10011.

FUNDAMENTAL DIAGRAM

Human Factors in the Fundamental Diagram**DAIHENG NI****CHAOQUN (ENZO) JIA***University of Massachusetts, Amherst***LINBO LI***Tongji University***HAIZHONG WANG***Oregon State University*

Some observations are made on the fundamental diagram of freeway traffic, among which are three influence regions, three types of transition around capacity, and capacity drop phenomenon. This research aspires to interpret these observations from a human factors perspective in traffic flow theory. Of particular interest are the following human factors: drivers' choice of desired speeds, perceived effective vehicle lengths, perception–reaction times, and aggressiveness. It appears that all four factors are involved in characterizing traffic flow, though they play different roles in the fundamental diagram. Different combinations of these factors give rise to the above-mentioned influence regions, transition around capacity, and capacity drop. Of critical importance in determining the transition around capacity and capacity drop is drivers' aggressiveness, a factor that has long been overlooked in the past. This paper provides a detailed account of where it comes from and how it influences the fundamental diagram.

SOME FIELD OBSERVATIONS

Figure 1 illustrates an empirical fundamental diagram observed on Georgia 400, a toll road in Atlanta, Georgia. Each point in the “cloud” represents a 5-min aggregation of the original 20-s observations, and the diagram consists of 1 year's worth of data.

Observation 1. Three Influence Regions

Strikingly in this diagram are the following regions. One is the area bounded by lines L_1 and L_2 just like a beam of light emitting from the origin O ; the other is the area bounded by lines L_3 and L_4 like another beam of light emitting from point D , which represents jam density k_j . The area $Q_1Q_2Q_3Q_4$ indicates a third region where the two beams of light interfere.

Obviously, area L_1OL_2 indicates free-flow conditions, where flow is dictated by free-flow speed; area L_3DL_4 consists of congested conditions, where flow is mainly constrained by drivers' perception–reaction capabilities. Influenced by both free-flow and congested regions, area $Q_1Q_2Q_3Q_4$ constitutes a capacity region, since this is where the capacity is most likely to be found.

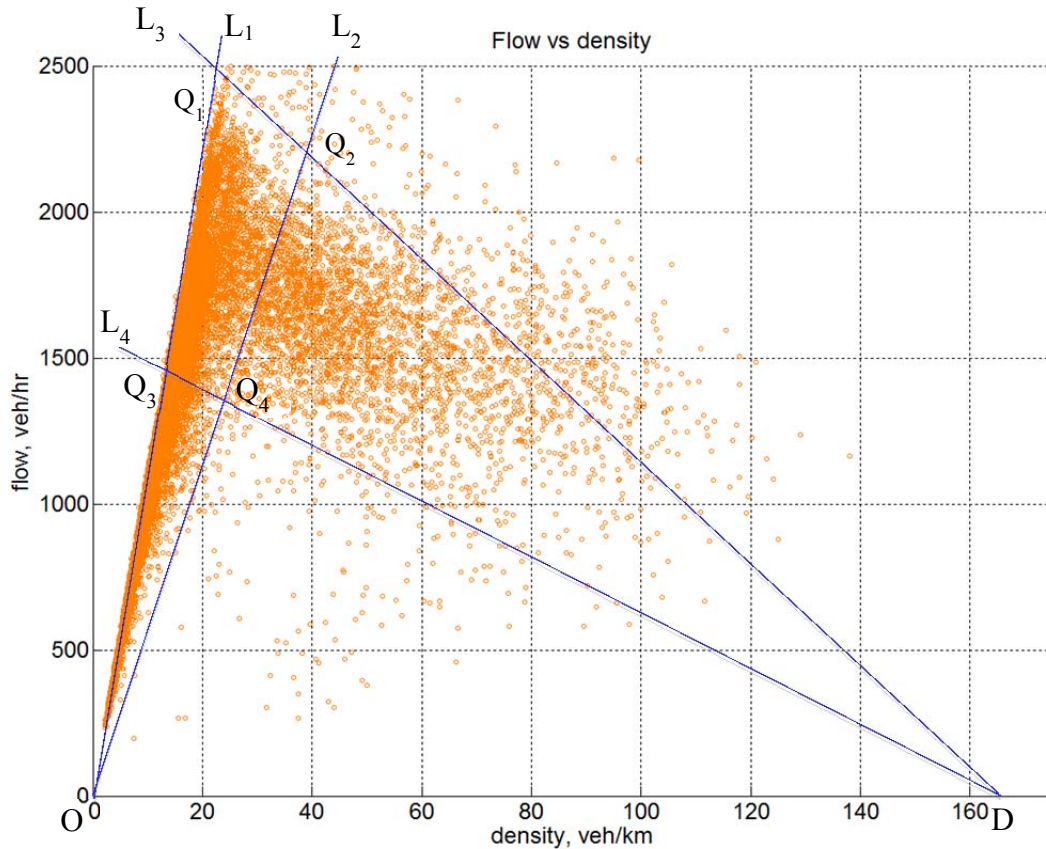


FIGURE 1 Fundamental diagram resulting from field observations.

It is intriguing to explore what factors are acting behind that drive the shaping of the diagram. Of key interests here are capacity condition (magnitude and location), the transition of flow–density relationship around capacity, and whether or not there is capacity drop phenomenon.

Observation 2. Three Types of Transition in Flow–Density Relationship

Observed in the field are the following three types of transition regarding flow–density relationship around capacity:

- Skewed-parabola (e.g., Curve 1 in Figure 2),
- Triangular (e.g., Curve 2 in Figure 2), and
- Reverse-lambda (e.g., Curve 3 in Figure 2).

Where a reverse-lambda transition is most likely found in inner-lane traffic, a triangular transition is often resulted in middle-lane traffic or by aggregating over all lanes, and a skewed-parabola transition is typical in outer-lane traffic. These empirical observations have been reported by Koshi et al. (1), Banks (2), and others.

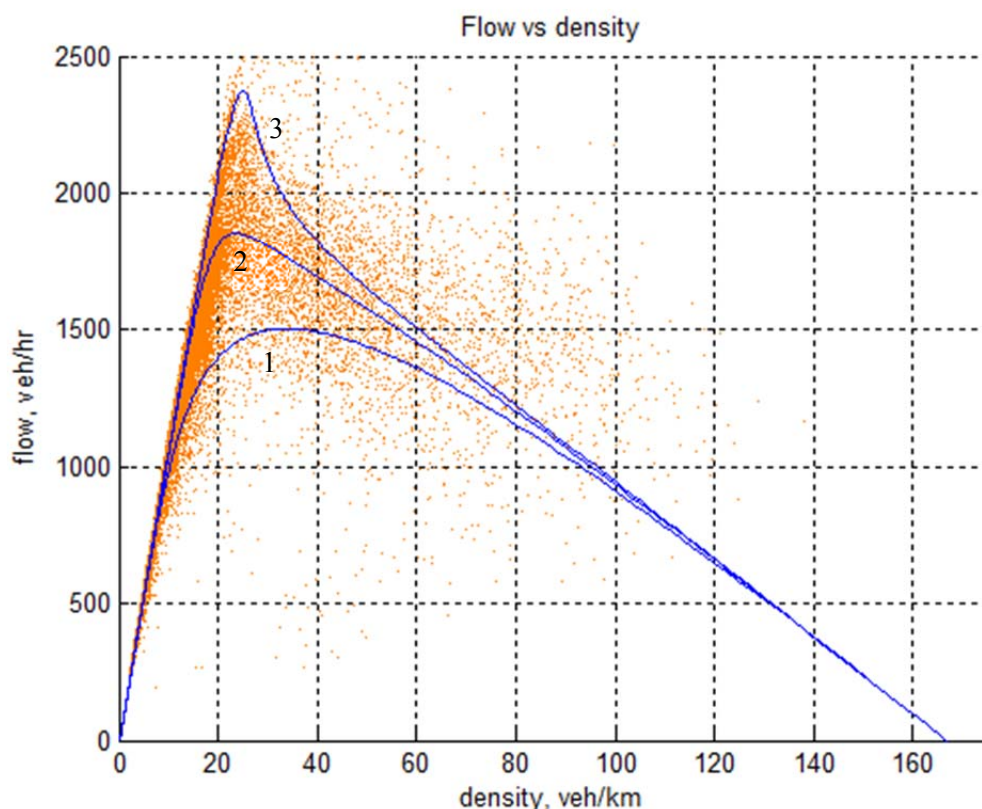


FIGURE 2 Flow–density relationships fitted to empirical data.

The authors have models that capture the skewed-parabola flow–density relationship as well as a few other models that represent the triangular flow–density relationship. It is of great interest to understand what gives rise to these types of flow–density relationships. In particular, what causes the reverse-lambda transition and how to model it?

Observation 3. Capacity Drop

Figure 3 is another fundamental diagram generated from GA400, but this figure includes only 1 day’s worth of data on a weekday (see the “cloud” in the background). Consisting of observations of flow and density in the field, each data point represents the operating condition of traffic over this observation period. A time development of operating conditions is plotted on top of the “cloud” with numbers indicating the temporal order of the observations of roughly 10 min apart. Note that the peak flow occurs at around 7:30 a.m., which is numbered 14 in the figure.

Following the time development, the temporal evolution of traffic conditions around capacity seems to suggest a pattern sketched on the right of the figure. First, there is a loading process early in the morning as the road starts to empty and traffic demand increases over time. This is sketched as Curve 1 in the figure where flow increases almost linearly with density all

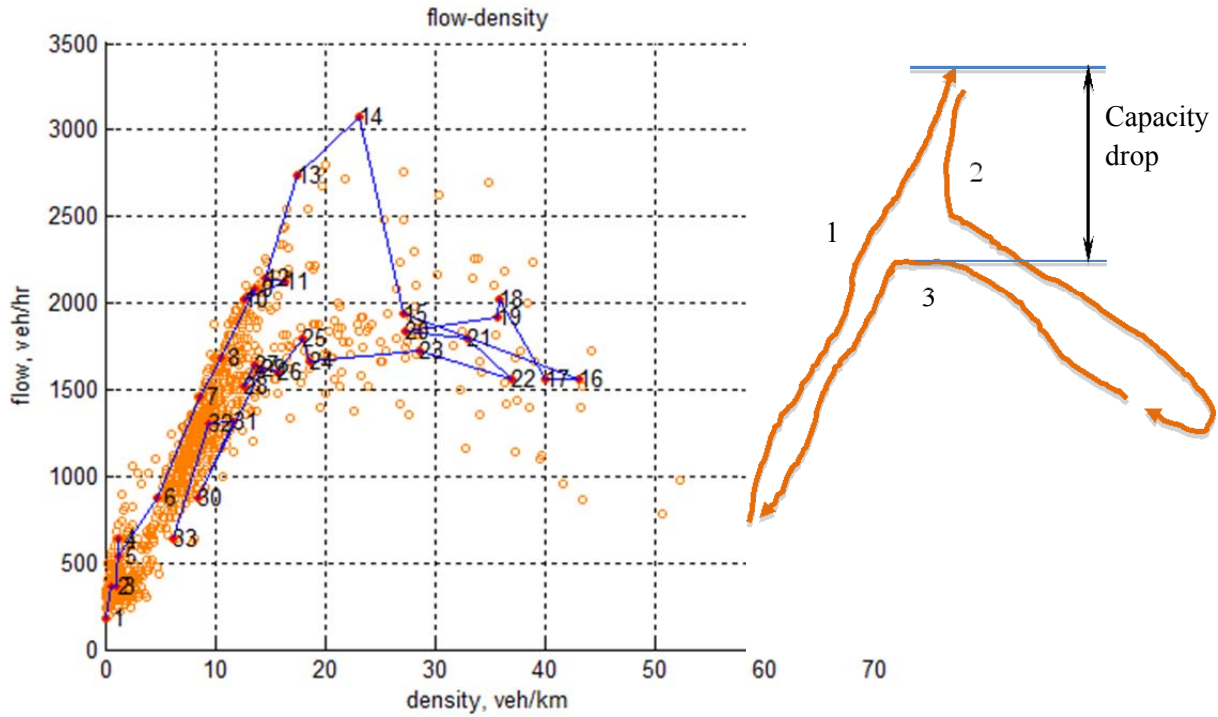


FIGURE 3 Time series plot of loading and unloading process.

the way up to the peak. As demand continues increasing beyond the peak, any disturbance may result in traffic breakdown and hence congestion. This corresponds to Curve 2 in the figure. As demand decreases, traffic begins to unload where density decreases, and flow increases as in the first half of Curve 3. However, the peak of the unloading curve is much less than that of the loading curve, and their difference constitutes a capacity drop phenomenon. With demand continually decreasing, both flow and density drop, as depicted in the second half of Curve 3. This concludes a cycle of traffic operation in a typical day. Similar findings of traffic loading, unloading, and capacity drop have been reported in Banks (3), Saberi and Mahmassani (4), and Chamberlayne et al. (5). Interested readers are motivated to investigate whether it is possible to capture such a phenomenon in traffic flow modelling.

With these observations and questions, the objective of this paper is to explore factors acting behind a fundamental diagram that give rise to these phenomena. The remainder of the paper is organized as follows. First, the authors start with examining human factors in car following and trace their way into macroscopic representation of traffic flow: the fundamental diagram. Next, they analyze the roles that these human factors play in fundamental diagram, in particular, which factor controls what part of the diagram. This is followed by revisiting empirical data and checking if field observations agree with the above analysis. To facilitate application of findings of this research, a method is formulated to roughly estimate human factors parameters from traffic flow data without requiring vehicle trajectory data. Lastly, findings are summarized and conclusions are drawn.

RELATING HUMAN FACTORS TO THE FUNDAMENTAL DIAGRAM

In essence, traffic flow is the result of action and interaction of many vehicles in a driving environment. The microscopic view of traffic flow is the result of human factors such as drivers' speed choice, perception–reaction time, acceleration and deceleration, and safety rules. On the other hand, the macroscopic view of traffic flow is typically reflected in a fundamental diagram that is characterized by macroscopic measures such as capacity, free-flow speed, and jam density. Micro-macro coupling is the key to unlocking fundamental diagram. In this paper, the authors are particularly interested in how human factors are related to highway capacity and traffic congestion and, in return, how the discovery would shed light on traffic flow modelling.

Microscopic View: Human Factors

Stopping Sight Distance

Traffic engineering books (6–8) have it as a standard that safe stopping sight distance (*SSD*) has to be maintained at any point on a roadway. The *SSD* is the sum of distance traveled during perception–reaction time τ and braking distance:

$$SSD = \tau v + \frac{v^2}{2b}$$

where v is vehicle speed and b is deceleration rate.

Good Driving Rule

Similarly, Forbes and Simpson (9) and equivalently Pipes (10) applied a “good driving rule” (GDR) to car following and stipulated that the minimum spacing s^* between a leading vehicle j and a following vehicle i be

$$s_{ij}^* = \tau_i \dot{x}_i + l_j$$

Where l is effective vehicle length (i.e., the actual vehicle length plus some buffer space at both ends).

Safe Driving Rule

Gipps (11) took it further by considering more dynamic and conservative situations between vehicles with speeds \dot{x}_i and \dot{x}_j and deceleration rates b_i and B_j as follows. At any time, vehicle i must leave enough spacing ahead such that, when vehicle j suddenly brakes, the spacing should allow vehicle i to stop safely behind vehicle j after a perception–reaction process and a braking process. A simplified expression of this “safe driving rule” (SDR) can be formulated as

$$s_{ij}^* = \frac{\dot{x}_i^2}{2b_i} - \frac{\dot{x}_j^2}{2B_j} + \tau_i \dot{x}_i + l_j$$

If the vehicles travel at the same speed, $\dot{x}_i = \dot{x}_j$, the above equation reduces to the following form (12, 13) (Chapter 4):

$$s_{ij}^* = \gamma_i \dot{x}_i^2 + \tau_i \dot{x}_i + l_j$$

where the coefficient of the nonlinear term γ_i is

$$\gamma_i = \frac{1}{2} \left(\frac{1}{b_i} - \frac{1}{B_j} \right)$$

Unlike perception–reaction time τ , the physical meaning of γ was not defined, but the coefficient was generally regarded as positive. An empirical value for γ would be 0.023 s²/ft (about 0.075 s²/m) (12).

Aggressive Driving Rule

A less conservative consideration of γ_i (14) may be the result of choosing proper values for b_i and B_j . For example, it is recognized that b_i actually represents the deceleration rate at which driver i believes that he or she is capable of applying in an emergency. Of particular interest is the possibility that b_i might be greater in magnitude than B_j , which represents driver i 's estimate of the emergency deceleration that is most likely to be applied by driver j . Consequently, γ_i may take negative values: $\gamma_i < 0$, which gives rise to an “aggressive driving rule.” Rather than creating an extra safety buffer as in the SDR, the negative driving rule even shortens the safe distance stipulated by the GDR.

In sum, when $\gamma_i = 0$ as in the Forbes model, the GDR is resulted; when $\gamma_i > 0$ as in the Gipps model, the SDR is resulted since the driver leaves extra room; when $\gamma_i < 0$, the aggressive driving rule (ADR) is resulted in that the driver tends to follow dangerously close. Hence, it is reasonable to associate γ_i to the driver's aggressiveness, since it represents the driver's willingness to run risk in trade of speed gain.

Longitudinal Control Model

The ADR alone was not modeled. However, Ni (15) recently incorporated the concept of aggressiveness into the Longitudinal Control Model (LCM):

$$\ddot{x}_i(t + \tau_i) = A_i \left[1 - \left(\frac{\dot{x}_i}{v_i} \right) - e^{1 - \frac{s_{ij}}{s_{ij}^*}} \right]$$

where v_i is driver i 's desired speed and A_i is the maximum acceleration desired by driver i when starting from standing still. Note that $s_{ij}^* = \gamma_i \dot{x}_i^2 + \tau_i \dot{x}_i + l_j$ and aggressiveness γ_i is allowed to be positive, zero, and negative representing a spectrum of driving rules.

The discussion on traffic flow models thus far has involved a few human factors parameters including aggressiveness γ_i , perception–reaction time τ_i , desired speed v_i , and effective vehicle length l_i . It is interesting to examine how these parameters manifest themselves

in the fundamental diagram, where individual behaviors are aggregated to exhibit collective properties of traffic flow.

Macroscopic View: The Fundamental Diagram

Characterizing the fundamental diagram are flow-speed-density relationships. Under steady-state condition (i.e., traffic state does not change in time), the GDR gives rise to the following flow q and density k relationship:

$$q = \frac{1}{\tau} - \frac{l}{\tau}k$$

where τ and l are perception–reaction time τ_i and effective vehicle length l_j , respectively, aggregated over all vehicles. Similarly, the flow–density relationship implied by the SDR is given in parametric form:

$$q = kv \quad \text{and} \quad k = \frac{1}{\gamma v^2 + \tau v + l}$$

Where γ and l are γ_j and l_j , respectively, aggregated over all vehicles. The parametric flow–density relationship of LCM is

$$q = kv \quad \text{and} \quad k = \frac{1}{(\gamma v^2 + \tau v + l)[1 - \ln\left(1 - \frac{v}{v_f}\right)]}$$

where free-flow speed v_f is desired speed v_i aggregated over all vehicles.

INFLUENCE OF HUMAN FACTORS ON THE FUNDAMENTAL DIAGRAM

Desired Speed v_i

The foremost of the set of human factors parameters is desired speed v_i . When traffic is light and actual spacing s_{ij} is greater than desired spacing s_{ij}^* , drivers are insensitive to neighboring vehicles and are free to choose their desired speed v_i . This is what the SDR, GDR, and LCM stipulate:

$$\text{SDR: } \dot{x}_i = v_i \quad \text{when} \quad s_{ij} > s_{ij}^* = \gamma_i \dot{x}_i^2 + \tau_i \dot{x}_i + l_j$$

$$\text{GDR: } \dot{x}_i = v_i \quad \text{when} \quad s_{ij} > s_{ij}^* = \tau_i \dot{x}_i + l_j$$

$$\text{LCM: } \dot{x}_i = v_i \left(1 - e^{-\frac{s_{ij}}{s_{ij}^*}}\right) \quad \text{for all } s_{ij}, \quad \text{where } \dot{x}_i \sim v_i \quad \text{when } s_{ij} > s_{ij}^*.$$

When aggregated over drivers, desired speed becomes free-flow speed: $v_i \rightarrow v_f$. Therefore, all three models agree that desired speed dictates free-flow speed which, in turn, determines traffic conditions in the free-flow region of Figure 1. As Forbes (16) pointed out, flow q increases linearly with density k without being constrained by leading vehicles. This effect is illustrated in the top left subplot of Figure 4, which is generated using LCM. In this plot, free-flow speed varies, while the other three parameters are held constant as indicated. Units are in metric system.

The capacity condition of SDR is found at

$$q_m = \frac{1}{2\sqrt{\gamma l} + \tau} \quad v_m = \sqrt{\frac{\gamma}{l}} \quad k_m = \frac{l}{\gamma^2 + \tau\sqrt{\gamma l} + l^2}$$

Similarly, GDR yields

$$q_m = \frac{v_f}{\tau v_f + l} \quad v_m = v_f \quad k_m = \frac{1}{\tau v_f + l}$$

LCM does not yield a closed form of capacity but can be solved numerically. It is interesting to note that, in SDR, capacity is irrelevant to free-flow speed, while both GDR and LCM predict that capacity is a function of free-flow speed. The effect is also illustrated in the top left subplot of Figure 4.

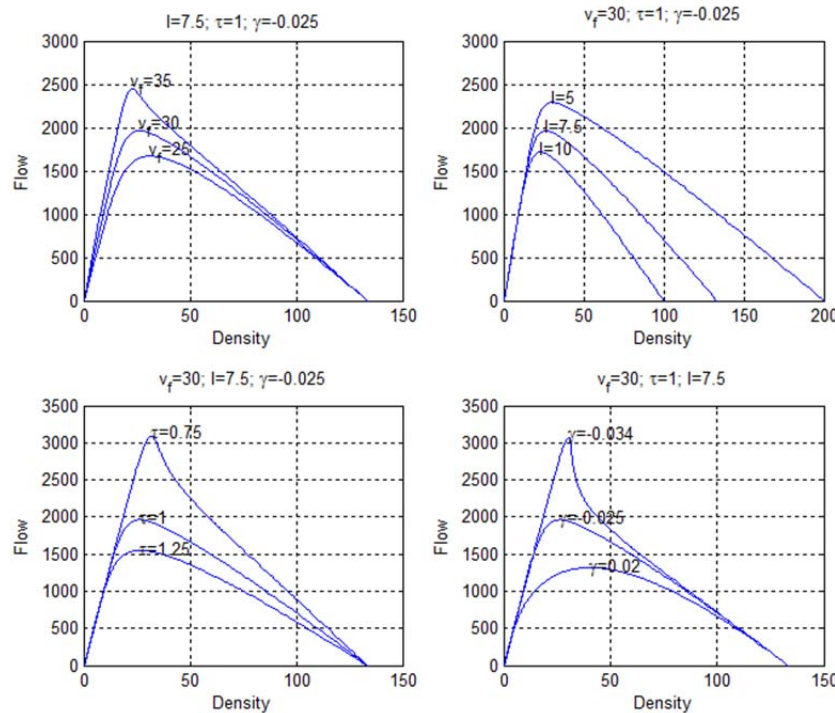


FIGURE 4 Influence of human factors on the fundamental diagram.

Effective Vehicle Length

Effective vehicle length l_i is the longitudinal space that a driver sees himself or herself representing in the traffic. It is typically regarded as the physical vehicle length plus some buffer spaces at both ends. It is also viewed as the bumper-to-bumper spacing when traffic is jammed. When aggregated over vehicles, effective vehicle length is the reciprocal of jam density: $l_i \rightarrow l = 1/k_j$. Unlike free-flow speed, which is accurately visible in field data, jam density is seldom revealed (see Figures 1 to 3). However, it can be deduced by following the trend of the congested region or determined from a reasonable range, for example, 100 to 200 veh/km based on an effective vehicle length of 5 to 10 m. The influence of l is illustrated in the top right subplot of Figure 4, where l is varying while others parameters are held constant.

In stop-and-go traffic around jammed conditions, the stopping or moving of vehicles propagates backward, progressively forming a kinematic wave at jam density ω_j . Since drivers react strictly according to driving rules, the speed of the kinematic wave at jam density ω_j can be determined as

$$\text{SDR: } \omega_j = \left. \frac{dq}{dk} \right|_{k=k_j} = \left(v - \frac{\gamma v^2 + \tau v + l}{2\gamma v + \tau} \right)_{v=0} = -\frac{l}{\tau}$$

$$\text{GDR: } \omega_j = \left. \frac{dq}{dk} \right|_{k=k_j} = -\frac{l}{\tau}$$

$$\text{LCM: } \omega_j = \left. \frac{dq}{dk} \right|_{k=k_j} = -\frac{l}{\tau + l/v_f}$$

Apparently, effective vehicle length affects ω_j , and the effect is also illustrated in the top right subplot of Figure 4. Assume effective vehicle length of 6 m, perception–reaction time of 1 s, and free-flow speed of 30 m/s (or 108 km/h), SDR and GDR suggest an ω_j of -21.6 km/h, LCM results in -18 km/h. These results are in agreement with field observations as well as with literature [e.g., -23 km/h in Lighthill and Whitham (17) and -20 km/h in Del Castillo and Benítez (18)].

As discussed previously, effective vehicle length also affects capacity condition: its magnitude and location.

Perception–Reaction Time

Though free-flow speed appears insensitive to perception–reaction time τ_i , capacity is heavily influenced by this parameter as confirmed by the capacity condition above. Since τ_i represents how fast a driver responds to stimuli (e.g., braking light of the front vehicle), it mainly takes effect in relatively dense traffic, where drivers need to pay attention to surrounding vehicles, especially the leading one. Therefore, the congested region is also heavily influenced by perception–reaction time. A disturbance in traffic is almost sure to be picked up by following vehicles and gets propagated upstream. Hence, backward wave speed u has a close relationship with τ_i . As a boundary condition, the speed of a backward wave at jam density ω_j was given above. In general, the shorter the perception–reaction time (i.e., $\tau_i \downarrow$), the greater the backward wave speed in magnitude (i.e., $|u| \uparrow$), and hence the steeper (more negative) the congested branch of the corresponding flow–density relationship. The effects are illustrated in the bottom left subplot of Figure 4.

As the congested branch becomes steeper, it meets the free-flow branch at higher flow rates suggesting greater capacities. However, it seems that perception–reaction time is not the only parameter that dictates how the two branches meet and hence the underlying flow–density relationship.

Aggressiveness

It turns out that aggressiveness γ_i plays an even more important role in the shaping of fundamental diagram, more specifically, capacity condition (magnitude and location), the transition around capacity, and whether or not there is capacity drop phenomenon.

In the SDR, aggressiveness assumes positive values $\gamma_i > 0$, which corresponds to timid drivers who need extra room for increased safety. As drivers become more timid (i.e., $\gamma_i \uparrow$), they will need longer safe distances, and hence less flow can be sustained. This relationship is reflected in the capacity condition given above.

In addition, timid traffic (i.e., $\gamma_i > 0$) results in early termination of free-flow condition and early onset of congested condition, because drivers tend to respond to leading vehicles at lower density than there would be for normal (i.e., $\gamma = 0$) or aggressive (i.e., $\gamma < 0$) traffic. Meanwhile, the transition is smooth (i.e., a skewed parabola flow–density curve is resulted around capacity condition. See the bottom right subplot of Figure 4).

The GDR does not have an aggressiveness term, which can be viewed as $\gamma_i = 0$. Since aggressiveness is zero here, it does not play a role in the formula of capacity condition. Meanwhile, zero aggressiveness in the GDR always yields a triangular flow–density relationship, meaning discontinuity is introduced at capacity condition (q_m, k_m) where the free-flow condition (i.e., $q = vk$) and congested condition (i.e., $q = \frac{1}{\tau} - \frac{l}{\tau}k$) meet.

LCM does not pre-assume a value for γ_i which can be positive, zero, or negative depending on driver population. When traffic exhibits aggressive characteristics (i.e., $\gamma < 0$), LCM predicts that free-flow condition may sustain up to very high flow rates (e.g., over 3,000 veh/h) before giving in to congested condition. In addition, such a transition can be a drastic one in that flow drops significantly when the changeover takes place. It appears as though there were two capacities here. One results when loading traffic onto an empty highway. Flow increases linearly with density. As density continues increasing, flow keeps its momentum and rises up to levels that normally require drivers to observe safe distance. At this point, traffic seems to be oversaturated. Consequently, a little disturbance in traffic would result in fast deterioration in conditions. All of the sudden, traffic becomes packed, with a significant drop in speed and flow. Another capacity is typically resulted during an unloading process from congested condition. As traffic keeps recovering toward the free-flow side, no capacity jump is observed around the pervious transition point. Instead, the new transition takes place smoothly, as in Figure 3, which suggests another capacity that is lower than the previous one (i.e., a capacity drop phenomenon). This results in a reverse-lambda flow–density relationship that neither of the other two models captures, and the effect is illustrated in the bottom right subplot of Figure 4.

The above discussion of influence of human factors parameters is summarized in [Table 1](#).

WHAT DO EMPIRICAL DATA SAY?

Loading Processes at a Few Locations

Generated from highway data at a few locations worldwide, Figure 5 samples the temporal development of the loading processes on these highways, including the Autobahn in Germany, Amsterdam Ring Road, Highway 401 in Canada, I-4 in Orlando, and PeMs in California. First of all, it is clear that free-flow condition is rather linear across the board, suggesting that drivers are not sensitive to leading vehicles and are free to choose their desired speeds. Note that free-flow speeds differ on different highways suggesting that drivers' choice of desired speeds is location specific.

Remarkable in the figure is when the curves peak as the loading process proceeds. For instance, the free-flow curve of PeMs is only sustainable at a little over 1,000 veh/h, while those of Amsterdam and Autobahn maintain up to about 2,500 veh/h. Though unknown without field measurement, perception–reaction time can be roughly deduced from peak flow rates. For example, according to the GDR, average headway h is perception–reaction time plus a little buffer, $h = \tau + l/v_f$. PeMs data suggest a perception–reaction time of a little less than 3.0 s, while that of Amsterdam is roughly 1.4 s. Given that free-flow speed is about the same at the two locations, one would conclude that drivers at the two locations have about the same taste of desired speed but in general take differently time to react to leading vehicles.

Also striking in the figure is how the curves peak. More specifically, this concerns how free-flow condition gives in to congested condition. For example, PeMs and Amsterdam data seem to transition along a skewed parabola curve; I-4 data suggest a triangular changeover; Autobahn and Highway 401 data appear to experience a capacity drop when the congested region takes over (i.e., a reverse-lambda shape).

At the minimum, Figure 4 reveals that different locations give rise to different driver characteristics, which, in turn, result in different traffic flow characteristics. For example, the five sets of data suggest different free-flow speeds, different peak flow rates, different peak densities, and different transition curves. This poses a great challenge to traffic flow modeling, which should not only capture the underlying mechanism but also the wide ranges of variability.

TABLE 1 Influence of Human Factors on the Fundamental Diagram

Human Factor Parameters	v_i	l_i	τ_i	γ_i
Free-flow speed, v_f	Y	N	N	N
Capacity: magnitude, q_m	N Y Y	Y	Y	Y N Y
Capacity: location, k_m	N Y Y	Y	Y	Y N Y
Capacity: transition	N N Y	N	N N Y	Y
Capacity: drop	N N Y	N	N N Y	N N Y
Backward wave speed at jam, ω_j	N	Y	Y	N
Jam density, k_j	N	Y	N	N

NOTE: Y = has influence; N = no influence. A single letter denotes that the same comment applies to the three models. Three letters (e.g., N N Y) means the comments of SDR, GDR, and LCM, respectively.

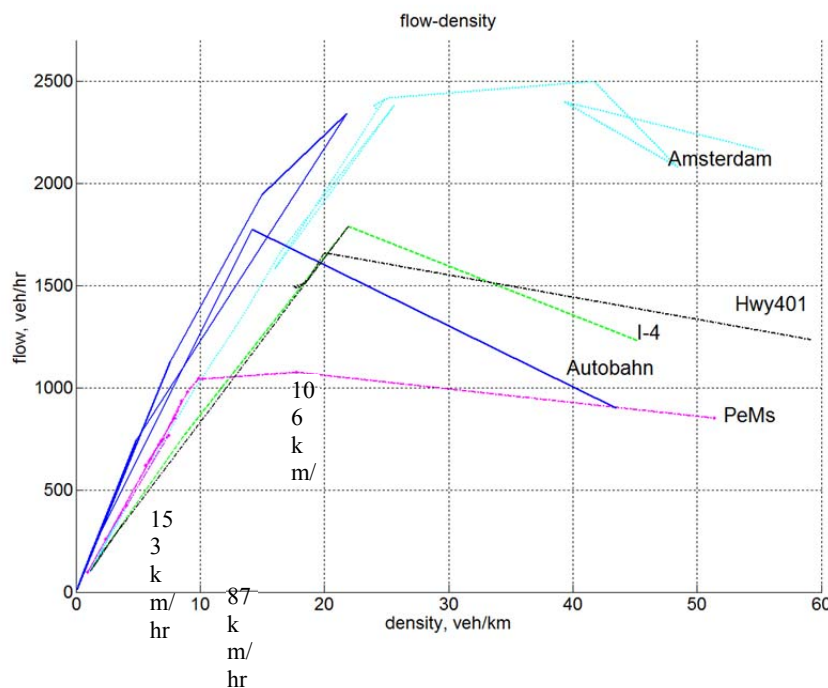


FIGURE 5 Temporal development of flow–density relationship on selected highways.

Fitting Models to Empirical Data

As a measure of how close modelling matches field observations, Figure 6 illustrates the result of fitting the three models to GA400 data with fitted parameters tabulated in Table 2.

The figure shows that the three models perform almost the same when traffic is light and severely congested. However, the difference is salient in the transition from free flow to congestion, particularly around capacity condition. With similar settings except for aggressiveness, the safe driving rule yields the lowest capacity, while LCM suggests the highest and also the closest to true capacity. This result tends to support the significance of aggressiveness in traffic flow modeling. Perhaps more prominent in the figure is the transition around capacity. The SDR with $\gamma = 0.023$ (meaning timid drivers) corresponds to a skewed parabola transition; GDR with $\gamma = 0$ (meaning normal drivers) introduces discontinuity that yields a triangular transition; LCM with $\gamma = -0.041$ produces a reverse-lambda transition suggesting a possible capacity drop.

In traffic flow modeling, it is important to faithfully reproduce free-flow, congested, and jammed conditions. It is even more important to capture capacity condition. As evidenced above, aggressiveness plays a differentiating role in capacity condition: magnitude, location, transition, and capacity drop. In addition, it is attractive for a model to represent a spectrum of transition around capacity such as skewed-parabola, triangular, and reverse-lambda, since they are all observed in the field. Note that the SDR captures only the first type, the GDR always yields the second type, and LCM is able to represent all three types (see Figure 4).

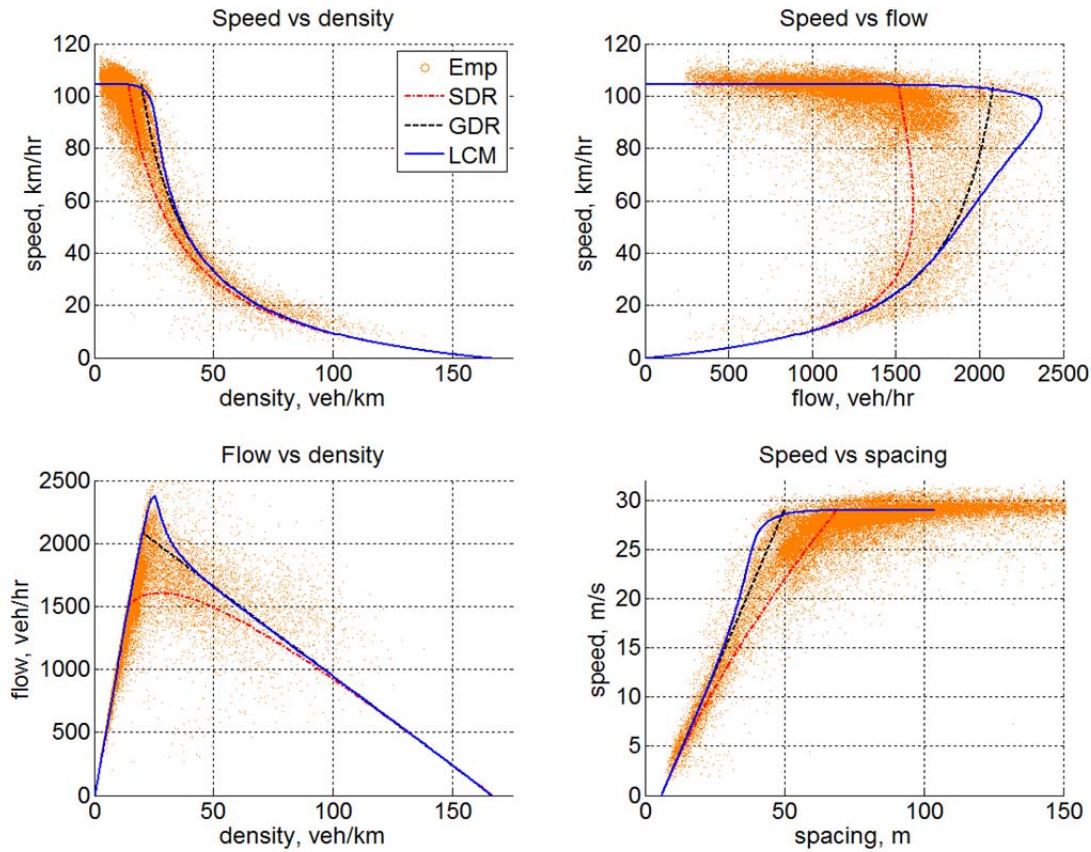


FIGURE 6 Three models fitted to empirical data.

TABLE 2 Human Factors Parameters Fitted to Empirical Data

Model	v_f (m/s)	τ (s)	l (m)	γ (s ² /m)
SDR	29	1.5	6	0.023
GDR	29	1.5	6	0
LCM	29	1.3	6	-0.041

Though Figure 6 only reveals a likely reverse-lambda transition around capacity and a possible capacity drop, these phenomena are confirmed in Figure 3 with temporal development of loading and unloading processes illustrated. Combined, Figures 3 and 6 seem to alert us to the possibility that loading and unloading processes may follow different speed–density relationships. For example, the loading process may be a reverse-lambda type, while the unloading is a skewed parabola or a triangle. If this is true, then two issues naturally follow. First, this observation seems to suggest that driver characteristics may be different during loading and unloading processes, and perhaps they may even be dynamic within a driving process. Such an effect is typically not captured in conventional modelling approaches. For example, perception–reaction time is normally treated as a constant in the authors’ models unless a stochastic approach is taken. Second, this observation may imply different types of flow–density

relationship for the same traffic during different processes. This requires that the underlying model be flexible enough to capture all of them without having to resort to another model. LCM is an example in case. Similarly, Chamberlayne et al. (5) produced the effect of capacity drop in INTEGRATION by varying vehicle acceleration behavior without having to introducing a discontinuous flow–density relationship.

ESTIMATION OF HUMAN FACTORS PARAMETERS

Involving measuring parameters, model calibration can be a time- and resource-consuming task, especially when sampling and aggregating a large population becomes necessary. This section intends to supplement such an endeavor with a quick, rough estimate by examining only traffic flow data that are readily available in most intelligent transportation systems. Assume one starts with an empirical diagram such as the clouds in Figures 1, 2, and 6, and somehow a flow–density relationship is sketched, as Curve 1 in Figure 7. Since field data typically show clear trend of free-flow condition, free-flow speed v_f can be found easily as the slope of the trend line in free-flow region. In addition, field data may also reveal a trend of congested condition. Following the trend, one may find jam density k_j as the intersection of the trend line and the horizontal axis.

Estimation of the remaining two human factors parameters, namely perception–reaction time τ and aggressiveness γ , necessitates two known points, A and B, on the curve. Though in the figure A appears to be the capacity and B possibly the reduced capacity, they don't have to be special in order to apply the estimation method. Since conditions at A(q_A, k_A, v_A) and B(q_B, k_B, v_B) are known, their average microscopic characteristics $A(h_A, s_A, v_A)$ and $B(h_B, s_B, v_B)$ can be calculated. Sketched on the right of the figure are trajectories of imaginary vehicles operating at conditions A and B. Using geometry of the trajectories and car-following rules, the following can be established:

$$\begin{cases} s_A = \gamma v_A^2 + \tau v_A + l \\ s_B = \gamma v_B^2 + \tau v_B + l \end{cases}$$

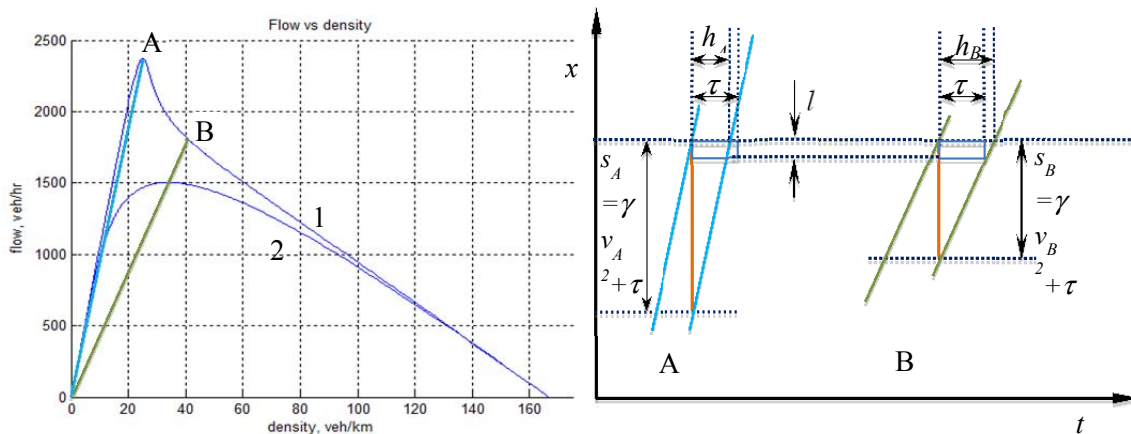


FIGURE 7 Estimation of human factors parameters.

Solving the above system of equations yields the following estimates:

$$\begin{cases} \gamma = \frac{(s_A v_B - s_B v_A) - l(v_B - v_A)}{v_A^2 v_B - v_A v_B^2} \\ \tau = \frac{-(s_A v_B^2 - s_B v_A^2) + l(v_B^2 - v_A^2)}{v_A^2 v_B - v_A v_B^2} \end{cases}$$

Unlike what is required by the SDR, the figure purposely shows that the headway of condition A, h_A , is shorter than perception–reaction time. This is to alert readers to the possibility that aggressive drivers may follow the leader closer than that stipulated by the GDR (i.e., $\gamma < 0$). Paradoxically, short headways are frequently observed in traffic, particularly in inner lanes, but car-following accidents are few. Harris (14) was probably right that this is so “not because the separations which drivers allow are necessarily safe but because the emergencies that reveal the danger are rare.”

CONCLUSIONS

This research examined the role of human factors such as aggressiveness, perception–reaction time, desired speed, and effective vehicle length in microscopic car following and traced their way into macroscopic traffic flow representation (i.e., fundamental diagram). Then, the influence of these human factors on the diagram was analyzed with reference to empirical observations in the field.

The findings of this research are summarized as follows. First, all four factors are actively involved in characterizing traffic flow, though they play different roles in fundamental diagram. More specifically, desired speed affects free-flow speed which, in turn, dictates free-flow condition. Effective length dictates jam density and affects backward wave speed at jam density. Perception–reaction time plays a significant role in virtually all aspects of fundamental diagram. In particular, perception–reaction time is critical in determining capacity (magnitude and location). No less than that of perception–reaction time, the role of aggressiveness has long been overlooked. Contrary to what is commonly believed, empirical evidence shows that drivers will follow at distances shorter than what is deemed as safe. Hence, aggressiveness is related to a driver’s willingness to tailgate in pursuit of speed gain. This research reveals that aggressiveness is the key to unlocking the mysteries in fundamental diagram such as reverse-lambda, flow-density relationship, and capacity drop.

In addition, empirical observations seem to suggest that driver characteristics may be different during loading and unloading processes, and perhaps they may even be dynamic within a driving process. Such an effect is typically not captured in conventional modeling approaches. For example, perception–reaction time is normally treated as a constant in the authors’ models unless a stochastic approach is taken. Meanwhile, empirical observations may imply different types flow–density relationship for the same traffic during different processes. This requires that the underlying model be flexible enough to capture all of them without having to resort to another model.

Moreover, this research provides a quick estimation method to roughly deduce human factors parameters from macroscopic traffic flow data without resorting to microscopic vehicle

trajectory data. However, this estimation method is intended as a supplement rather than a replacement to field measurements.

ACKNOWLEDGMENT

This research is supported in part by the University Transportation Center program.

REFERENCES

1. Koshi, M., M. Iwasaki, and I. Okhura. Some Findings and an Overview on Vehicular Flow Characteristics. *Proc., 8th International Symposium on Transportation and Traffic Flow Theory*, Toronto, Canada, 1983, pp. 403–426.
2. Banks, J. H. Freeway Speed-Flow-Concentration Relationships: More Evidence and Interpretations. In *Transportation Research Record 1225*, TRB, National Research Council, Washington, D.C., 1989, pp. 53–60.
3. Banks, J. H. Review of Empirical Research on Congested Freeway Flow. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1802*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 225–232.
4. Saberi, M., and H. Mahmassani. Capacity Drop Phenomena in Urban Networks. Presented at Midyear Meetings and Symposium on Innovations in Traffic Flow Theory, Fort Lauderdale, Fla., 2012.
5. Chamberlayne, E., H. Rakha, and D. Bish. Modeling the Capacity Drop Phenomenon at Freeway Bottlenecks Using the INTEGRATION Software. *Transportation Letters: The International Journal of Transportation Research*, Vol. 4, No. 4, 2012, pp. 227–242.
6. AASHTO. *A Policy on Geometric Design of Highways and Streets*, 6th ed., 2011.
7. Roess, R., E. Prassas, and W. McShane. *Traffic Engineering*, 4th ed., Prentice Hall, 2010.
8. Kilareski, W., S. Washburn, and F. Mannering. *Principles of Highway and Traffic Engineering*, 5th ed., John Wiley & Sons, 2012.
9. Forbes, T., and M. Simpson. Driver and Vehicle Response in Freeway Deceleration Waves. *Transportation Science*, Vol. 2, No. 1, 1968, pp. 77–104.
10. Pipes, L. A. An Operational Analysis of Traffic Dynamics. *Journal of Applied Physics*, Vol. 24, 1953, pp. 271–281.
11. Gipps, P. A Behavioral Car Following Model for Computer Simulation. *Transportation Research Part B*, Vol. 15, 1981, pp. 105–111.
12. *Highway Capacity Manual*. TRB, National Research Council, Washington, D.C., 1950.
13. *Revised Traffic Flow Theory Monograph*. Transportation Research Board. Washington, D.C., 2001.
14. Harris, A. J. Following Distances, Braking Capacity and the Probability of Danger of Collision Between Vehicles. *Austrian Road Research Board, Proceedings 2, Part 1*, pp. 496–512.
15. Ni, D. A Unified Perspective on Traffic Flow Theory, Part I: The Field Theory. *Applied Mathematical Sciences*, Vol. 7, No. 39, 2013, pp. 1929–1946.
16. Forbes, T. Human Factors Considerations in Traffic Flow Theory. *Highway Research Record 15*, HRB, National Research Council, Washington, D.C., 1963, pp. 60–66.
17. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves: A Theory of Traffic Flow on Long Crowded Roads. *Proc. of the Royal Society, Series A*, Vol. 229, 1955, pp. 317–345.
18. Del Castillo, J. M., and F. G. Benítez. On the Functional Form of the Speed-Density Relationship—I: General Theory. *Transportation Research Part B: Methodological*, Vol. 29, No. 5, 1995, pp. 373–389.

Traffic Flow Models

TRAFFIC FLOW MODELS

Congestion Scenario-Based Vehicle Classification Detection Models Based on Traffic Flow Characteristics and Observed Event Data

HENG WEI

QINGYI AI

HAO LIU

University of Cincinnati

ZHIXIA LI

University of Wisconsin–Madison

HAIZHONG WANG

Oregon State University

While the existing applied length-based vehicle classification model has been to estimate vehicle lengths accurately with dual-loop traffic monitoring station data under the free traffic condition, it produces considerable errors against congested traffic. In this study, both ground-truth vehicle trajectory and simultaneous loop event data are used to characterize the impact of congested traffic on vehicle classification. Eight scenarios are synthesized to define the vehicles' stopping locations over two single loops of the dual-loop station. Under the synchronized traffic flow, acceleration or deceleration is considered in the new developed Vehicle Classification under Synchronized Traffic model (VC-Sync model) to reflect the speed variation between loops. As a result, the error of the vehicle classification is reduced from 33.5% to 6.7%, compared to the existing applied model. Under the stop-and-go traffic condition, a stop-on-both-loops-only (SBL) was developed along with the VC-Sync model to simplify the complexity of the congested traffic situation in vehicle length estimation. Using the SBL model reduces the error from 235% to 17.1%, compared to the existing applied model. Capability of identifying traffic phases is a critical prerequisite to applying the new vehicle classification models under congestions. Therefore, an innovative method for identifying the traffic phases has been proposed based on the existing traffic stream models along with the new findings of the authors' empirical data analysis. As a result, a heuristic traffic phase identification model has been developed and successfully applied in the case study for evaluating the new length-based vehicle classification models with dual-loop data.

INTRODUCTION

This paper presents a scenario-based vehicle classification modeling method to estimate vehicle length via revealing possible scenarios of congested traffic impact on accuracy of vehicle length detection at a dual-loop station. The modeling effort addresses two issues: (1) identifying a sound solution to the problem of distinguishing congestion conditions that could be measured by loop data based on traffic flow characteristics and new findings resulting from analysis of the video-based vehicular event data, and (2) developing scenario-based models for improving

vehicle length estimation under congested traffic flows with evaluation of its improved accuracy by comparing the results with the existing applied model.

A dual-loop detector consists of two single-loop detectors placed with a fixed distance between single loops (e.g., 20 ft or 6 m), as shown in Figure 1. A vehicle can be detected by the dual-loop detector as electrical pulses of current are deduced in the loops when the vehicle enters and leaves the loop detection area. Each event of the electrical pulse is recorded as a timestamp. Normally, four timestamps, t_1 , t_2 , t_3 , and t_4 are recorded when a vehicle is operating through the loop detector area, as illustrated by Figure 1. This feature enables measuring traffic speed over the detection area, which is one of the key factors in estimating the vehicle length. Vehicle types are then identified in three or four “bins” based on the detected vehicle lengths.

In the existing applied vehicle classification model (which was then proven to be good for free traffic flow), no variation of a vehicle’s speed on both single loops is assumed (1). The existing model is described as follows:

$$speed = \frac{D}{t} \quad (1)$$

$$vehicle_length = speed \times \frac{OnT_2 + OnT_1}{2} - loop_length \quad (2)$$

where

$$\begin{aligned} D &= \text{distance between two loops (ft);} \\ t &= t_3 - t_1; \\ OnT_1 &= t_2 - t_1; \\ OnT_2 &= t_4 - t_3; \text{ and} \\ t_1, t_2, t_3, \text{ and } t_4 &= \text{time stamps when a vehicle enters or leaves the upstream loop (M loop) or} \\ &\quad \text{downstream loop (S loop) (Figure 1).} \end{aligned}$$

Under congested traffic, however, a vehicle’s speed changes frequently and even fiercely as it is traveling through the loops. In order to improve the accuracy of the vehicle length estimation against congested traffic, the authors extracted the ground-truth vehicle event data from video by using the software VEVID (2), which was finally compiled into high-resolution vehicular trajectory data. Meanwhile, simultaneous event data are derived from the dual-loop data. The sampling dual-loop station is located in the freeway I-71/I-74 in Columbus, Ohio (3). Both datasets were used to define scenarios of vehicles’ maneuvers as traversing through the loops and model the traffic conditions based on applied traffic stream characteristics and relevant theories. Finally, new models suitable for congested flows were developed and evaluated with the ground-truth data.

LITERATURE REVIEW

Greenshields (4) firstly proposed the traffic stream theory addressing the relationships among flow rate, speed, and density, in which speed and density are assumed to be linearly correlated.

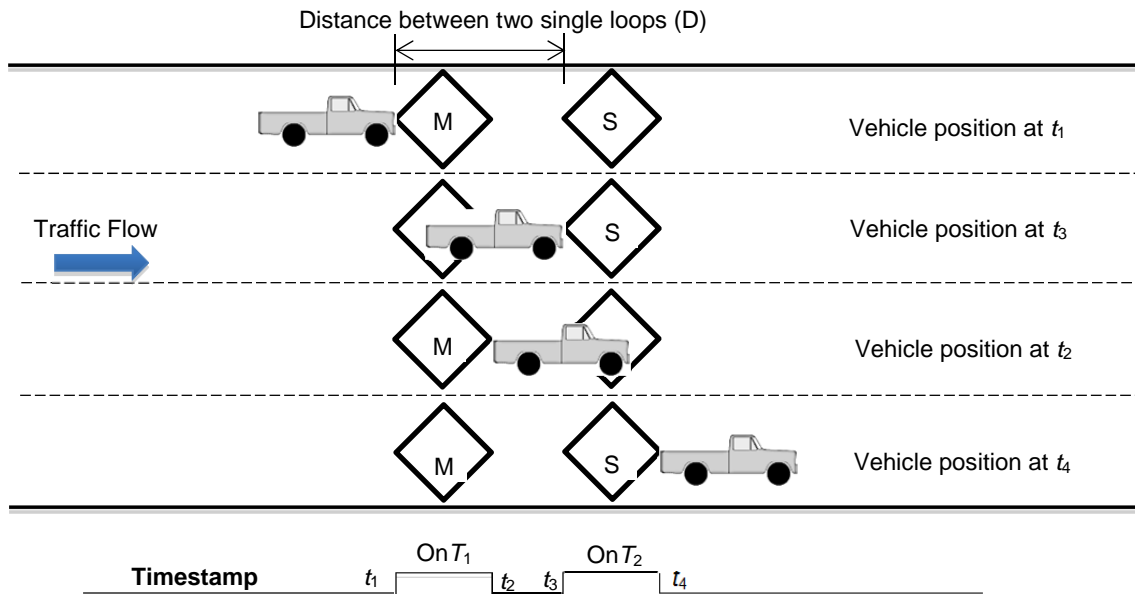


FIGURE 1 Layout of a dual-loop detector on highway.

Greenberg (5) revised the model of the speed and density to fit a logarithmic curve, based on a hydrodynamic analogy and assumption regarding the traffic flow as a perfect fluid and one-dimensional compressible flow. Underwood (6) used exponential expression for such a model. Researchers have disclosed the discontinuities of the relationships between traffic variables. Edie (7) quantified the linear relationship between density and the logarithm of velocity above the optimum velocity for uncongested traffic and velocity and the logarithm of spacing (the inverse of density) for congested traffic. Multiple curves are often applied to depict the discontinuities. For instance, Koshi (8) proposed a reverse lambda shape to describe the flow–density relationship. May (9) developed the “two-regime” models to describe the relationship of flow and density. Hall (10) proposed an inverted-V shape to represent the flow-occupancy relationship. Polus and Pollatschek (11) proposed three regimes of traffic flows (free, dense, and unstable flows), and traffic breakdown was explained as the change from dense flow to unstable flow.

Kerner and Konhäuser (12) and Kerner and Klenov (13) defined traffic flows in three categories: free flow, synchronized flow, and stop-and-go flow. The free flow has high travel speed and low traffic volume and density. The congested traffic flow is further classified into synchronized flow (S) and wide moving jam (J). The synchronized flow has relative low speed and high volume and density. A wide moving jam is a moving jam that maintains the mean velocity of the downstream front of the jam as the jam propagates. They also disclosed the double Z-characteristic shape for relating speed and density. The empirical double Z-characteristic shape is used to depict the phase transitions between two different phases. $F \rightarrow S$ (free flow to synchronized flow) and $S \rightarrow J$ (synchronized flow to jam flow) transitions can be illustrated by a double Z shape (or termed Z-characteristic) for the $F \rightarrow S \rightarrow J$ (free to synchronized to jam conditions) transitions. The double Z-characteristic consists of a Z-characteristic for an $F \rightarrow S$ transition and a Z-characteristic for an $S \rightarrow J$ transition, as well as the phases associated with the critical speeds required for the phase transitions. The synchronized traffic defined by Kerner is also described as the traffic oscillation by other researchers (14–18). Treiber and

Kesting (19) studied the convective instability in congested traffic flow, and they classified congested traffic flow into five classes according to the stability that lead to significantly different sets of traffic patterns (20).

It is necessary to determine what traffic variables and thresholds of the selected traffic variables will be used to describe the traffic phases and identify the transitions between them. Habib-Mattar et al. (21) found out that the congestion would occur if the situation, where the speed is less than 37 mph and the density is greater than 64 vpmpl, lasts at least 5 min. Chow et al.'s study (22) indicates that if the speed drop is greater than 5 mph during a 5-min period, then the traffic flow is at the congestion situation. Lorenz and Elefteriadou (23) defined a traffic breakdown as the traffic condition in which the average speed of all lanes on a highway section decreases to below 90 km/h for at least a 15-min period, and then Elefteriadou and Lertworawanich (24) changed the speed threshold to below 80 km/h. On the other hand, other studies indicated that speed alone is insufficient to ensure the identification of congestion. Congestion may not be detected by the speed-based algorithm only, and "perhaps the optimal speed thresholds are different above a certain occupancy threshold" (25). Zhang et al. (26) used four features to characterize an oscillatory traffic pattern: the occurrence of oscillation, the offset of the oscillation patterns different lanes, the oscillation period, and the oscillation amplitude in flow levels. They set the extreme jam density of 240 vpmpl, flow speed of 50 mph, and wave speed of 10 mph. Deng et al. (27) proposed a three-detector approach to identify traffic states using multiple data sources, including loop detector counts, AVI Bluetooth travel time readings, and GPS location samples. However, it is not always easy in practice to obtain the sensor data from all three sources for the traffic flow on a certain highway segment.

Since the event dual-loop data record individual vehicles' timestamps over the loops, they are usually applied in traffic analysis to derive traveling features of the vehicles (1, 28–31). The traffic parameters, such as traffic volume, speed, and occupancy or density can be extracted or calculated from the event dual-loop detector datasets, which further enable calculating vehicle lengths. The existing applied model of estimating vehicle lengths via dual-loop data (32) is based on the assumption that vehicles drive across the dual-loop detection area at a constant speed. The model has been validated well against light traffic. Under light traffic condition, vehicles operate at a relatively high and stable speed, which can be considered at a constant speed. According to Kerner's Three Phases theory, during uncongested traffic flow it is reasonable that vehicle speeds are regarded as constant. However, during congested traffic, especially stop-and-go traffic, vehicle speeds become very unstable and are not constant. When the existing model is used to estimate vehicle lengths, the accelerations and decelerations of vehicles will distort the outputs of the model. Accuracy of vehicle classification drops greatly under very congested traffic (33). It is reported that observed errors in truck misclassification ranged from 30% to 41% for off-peak hours, and from 33% to 55% for peak hours (32). Li (34) developed a method of Bayesian inference for vehicle speed and length estimation using dual-loop data. But the congested traffic flow features were not addressed in the method, and it was only tested using the traffic flow data with the average speed of 56 mph.

DATA COLLECTION

The selected dual-loop detector station, numbered as V1002, is located in the interstate freeway I-70/71 at West Mound Street downtown Columbus, and has six dual-loop detectors in both directions of the highway. A video camera was placed on the top of the Franklin County Juvenile Parking Garage that is close to the station to videotape the traffic flow on I-70/71 over the dual-loop detector station, as shown by [Figure 2](#).

Three-day traffic videotaping was conducted July 14–16, 2009. A total of 26 h of traffic video data were collected, including light traffic and congestion traffic flows. The concurrent event dual-loop data were obtained from the traffic management center (TMC) at the Ohio Department of Transportation (DOT). The event loop data is the raw data from the dual-loop station, which records the timestamps of each vehicle as it enters and leaves each loop. The scanning frequency of the loop is 60 Hz, that is, occupied status of a loop is automatically updated 60 times per second.

The ground-truth data used in this study is the vehicle trajectory data extracted from the collected traffic video footage. The software VEVID (2) was employed to extract the ground-truth vehicle trajectory data from the video.

A QSTARZ BT-Q1200 Ultra GPS Travel Recorder was adopted as the data logger to collect GPS data. The GPS travel data logger was equipped in a probe car running roundly along freeway segments of the I-70/I-71, which cover the selected station. The data logger can collect the probe vehicle's speed and location information by second. Some parameters that represent characteristics of very congested traffic can be derived from the statistical analysis of the collected GPS data, which include range of acceleration or deceleration rate and average minimum speed to maintain a vehicle's moving.

DISTINGUISHING TRAFFIC FLOW STATES OR CONDITIONS

Traffic Flow Condition Determined by Phase Representative Variables

Flow rate has been conventionally used as one of the measurable variables to depict the characteristics of the traffic flow in previous studies; however, application of the flow rate alone may be problematic to identifying the traffic conditions (or phases) when the length-based vehicle classification is practiced with dual-loop data. Firstly, any flow rate value may be

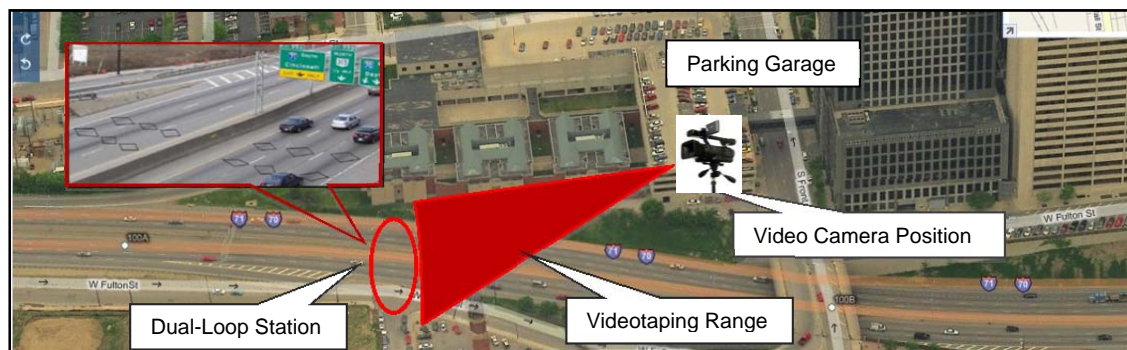


FIGURE 2 Video data collection and loop station at study site.

explained by two or more traffic phases (e.g., uncongested or congested traffic), which may cause a wrong identification of traffic condition. Secondly, the flow rate is an aggregated outcome from the dual-loop-based vehicle classification model and is supposed to be produced after the traffic phase is identified. That leads to an illogic procedure in practice. Timestamps and occupancies of a vehicle entering and leaving the loops are direct outputs of the loop data. Speed and density can be estimated as a mathematical function of the timestamps and occupancies. According to Kerner’s empirical double Z-characteristic shape (as shown in Figure 3), the speed and density are two variables that can be used to determine the boundaries of each traffic flow phase. The speed and density/occupancy are accordingly identified as the phase representative variables in this study.

In Kerner’s study (13) speed and density were applied to depict the empirical double Z-characteristic shape for the phase transitions between two different phases. The original Z-characteristic shape was enhanced and simplified in the study, as illustrated in Figure 3. It conceptually provides a profile of all the possible phases of traffic flows that could be justified by speed and density (or occupancy). Density can be estimated from the loop data by Equation 3 if the average vehicle length of the traffic flow for varying time of a day could be predetermined based on the historical traffic data.

$$K_i = \frac{1000 \times O_{cc}}{L_v + L_{eff}} \tag{3}$$

where

- K_i = density of the traffic flow (vpkmp/h) for time period i of a day,
- O_{cc} = loop occupancy measurement (%),
- L_v = average vehicle length (m), and
- L_{eff} = effective detector length (m).

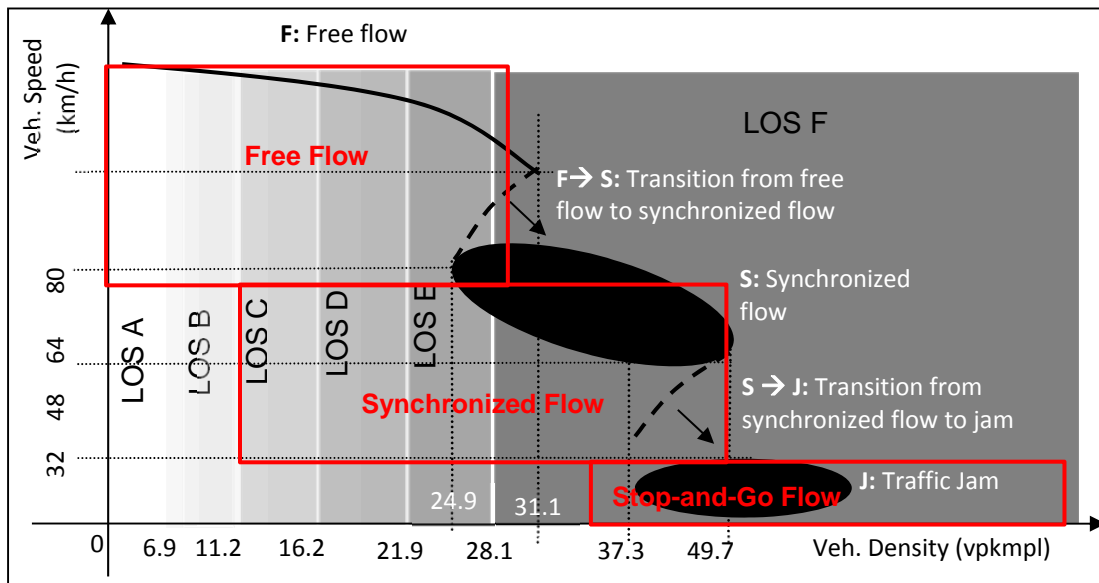


FIGURE 3 Classified traffic flow states (based on Kerner’s Z-curve and data in this study).

To simplify the procedure of the traffic condition identification, the F→S transition was merged into the free flow phase and S→J transition into the synchronized phase. Equation 4 was proposed to facilitate the development of a computing algorithm that will be used to determine the traffic flow phase $F(t_i)$ of any time period i .

$$F(t_i) = \begin{cases} \text{FF, IF } [u \geq 80 \ \& \ k \leq 28.1] \text{ OR IF } [\bar{v}(t) - \bar{v}(t+1) \leq \Delta v \ \& \ \text{var}(v) < v^*] \\ \text{SF, IF } [32 \leq u < 80 \ \& \ 11.2 \leq k \leq 49.7] \text{ OR IF } [(\bar{v}(t) - \bar{v}(t+1)) > \Delta v \text{ or } \text{var}(v) \geq v^*] \ \& \\ \quad (\overline{\text{occ}}(t) - \overline{\text{occ}}(t+1) \leq \Delta \text{occ}) \ \& \ (\overline{\text{occ}}(t) \leq \text{occ}^*) \\ \text{TJ, IF } [0 \leq u < 32 \ \& \ k \geq 31.1] \text{ OR IF } [(\bar{v}(t) - \bar{v}(t+1)) > \Delta v \text{ or } \text{var}(v) \geq v^*] \ \& \\ \quad (\overline{\text{occ}}(t) - \overline{\text{occ}}(t+1) > \Delta \text{occ}) \ \text{or} \ (\overline{\text{occ}}(t) > \text{occ}^*) \\ \text{SU, IF others} \end{cases} \quad (4)$$

where

- k = density (vpkmp1);
- u = speed (km/h);
- i = time period i ;
- FF = free-flow phase;
- SF = synchronized flow phase;
- TJ = traffic jam phase;
- SU = special or unreasonable case;
- t = a short period of time (5 min in this study);
- $\bar{v}(t)$ = the average speed in time interval t , km/h;
- $\bar{v}(t+1)$ = the average speed in the successive time interval $t+1$, km/h;
- $\text{var}(v)$ = the variation of all vehicles' speed during time interval t ;
- Δv = predefined threshold of spot speed difference in successive time intervals, km/h;
- v^* = predefined threshold of the speed variation range in successive time intervals, km/h;
- $\overline{\text{occ}}(t)$ = the average occupancy during time interval t ; and $\overline{\text{occ}}(t+1)$ = the average occupancy in the successive time interval $t+1$;
- Δocc = the predefined occupancy bandwidth during the time interval t ; and
- occ^* = the maximum average occupancy during the time interval t .

In this study, the percentage of types of vehicles and their average lengths are obtained from the sample dual-loop data at the dual-loop station V1002. The sample size is 13,722. The three-bin scheme standard adopted by Ohio DOT is used. The sample data indicates that the percentages of small vehicle (length ≤ 8.5 m), medium vehicle ($8.5 \text{ m} < \text{length} < 14.0$ m), and large vehicle (length ≥ 14.0 m) are 86%, 4%, and 10%, respectively. Their mean lengths are estimated as 5.0, 11.1, and 22.6 m, respectively. At V1002, L_{eff} is 2.6 m, and then, $L_v = 0.86 \times 5.0 + 0.04 \times 11.1 + 0.10 \times 22.6 = 7.0$ m. The assumed phase representative variables are evaluated against the real-world dual-loop data and the VEVID-based vehicular trajectory data. In light of the statistical analysis performed on the collected ground-truth and loop data, the thresholds of Δv is determined as 16.1 km/h, and v^* is determined as 127.7 km/h² (or the standard deviation is 11.3 km/h), Δocc is defined as 0.3, and occ^* is 0.35. To better understand the relationship between each defined traffic phase and the associated level of service (LOS), the LOS is overlaid in Figure 3 with corresponding density ranges as defined in the *Highway Capacity Manual 2010* (35).

MODELING SCENARIOS OF CONGESTED VEHICLE MANEUVERS OVER LOOPS

Under the synchronized traffic, vehicles speeds may change rapidly and frequently. In other words, a vehicle may drive over the upstream and downstream loops at different speeds as it increases or decreases its speed after leaving the upstream loop. Under this circumstance, the vehicle's acceleration or deceleration, which is not considered in the existing applied model, should not be ignored and is assumed to affect measurement of the vehicle length in great part. The characteristics of vehicle movement in the stop-and-go traffic flow are much different from the free or synchronized flow traffic. Vehicles are operating at a high, relatively constant speed under the free-flow traffic, and the free-flow traffic will transit to the synchronized traffic flow when the traffic speed drops significantly. The synchronized traffic flow will change into stop-and-go traffic when the traffic speed becomes very slow with more frequent acceleration or deceleration involved, and from time to time vehicles have to experience one or more stops. Under the stop-and-go traffic phase, a vehicle may stop within the dual-loop detection area for at least one time. The existing applied vehicle classification model produced more errors under the stop-and-go traffic, especially for large vehicles (see [Figure 4](#)), and the sample error even reaches 235%. It is observed from the comparison of the video-based vehicular event data and result from the existing applied model that the vehicle traveling features against stop-and-go traffic, such as acceleration or deceleration, and situation of vehicle stopping on loops, actually affect the estimation of vehicle lengths. An updated length-based vehicle classification model is therefore developed to improve the accuracy of vehicle length estimation under the stop-and-go traffic.

After careful analysis of synchronizing the ground-truth vehicular trajectory data and the dual-loop data, eight possible scenarios were synthesized based on possible stopping locations of the detected vehicles within the detection area, as illustrated by [Figure 5](#). Those eight scenarios are briefly described as follows.

Scenario 1. The vehicle drives across the dual-loop detection area without a stop, which is a typical synchronized flow feature.

Scenario 2. The vehicle stops merely on the M loop and then leaves the dual-loop detection area without another stop.

Scenario 3. The vehicle runs across the M loop and stops only on the S loop.

Scenario 4. The vehicle comes into the dual-loop detection area and stops only on both the M and S loops and then leaves the detection area without another stop.

Scenario 5. The vehicle stops on the M loop and then moves on and then stops on the S loop and finally leaves the detection area without another stop.

Scenario 6. The vehicle stops first on the M loop and then stops on both the M and S loops and finally leaves the detection area.

Scenario 7. The vehicle stops first on both of the M and S loops, and then stops only on S loop.

Scenario 8. The vehicle stops first only on the M loop and then stops on both of the M and S loop, and finally stops only on the S loop. Eventually the vehicle leaves the dual-loop detection area without another stop.

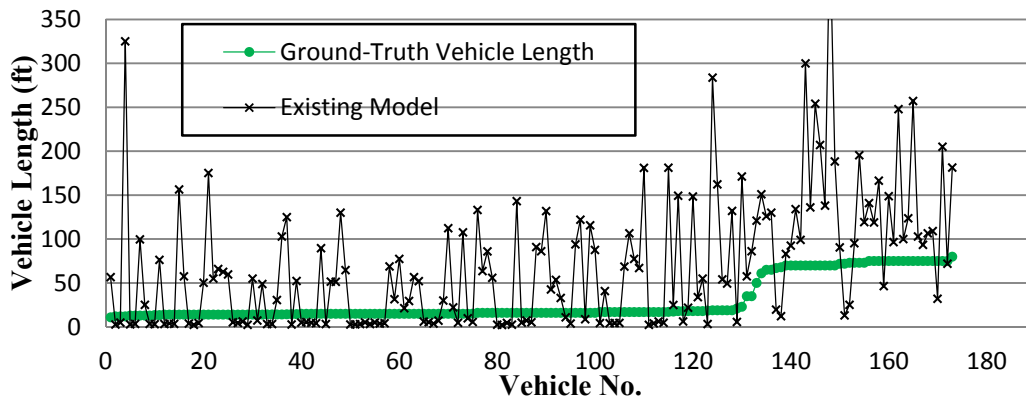


FIGURE 4 Vehicle length estimation of the existing applied model under stop-and-go traffic.

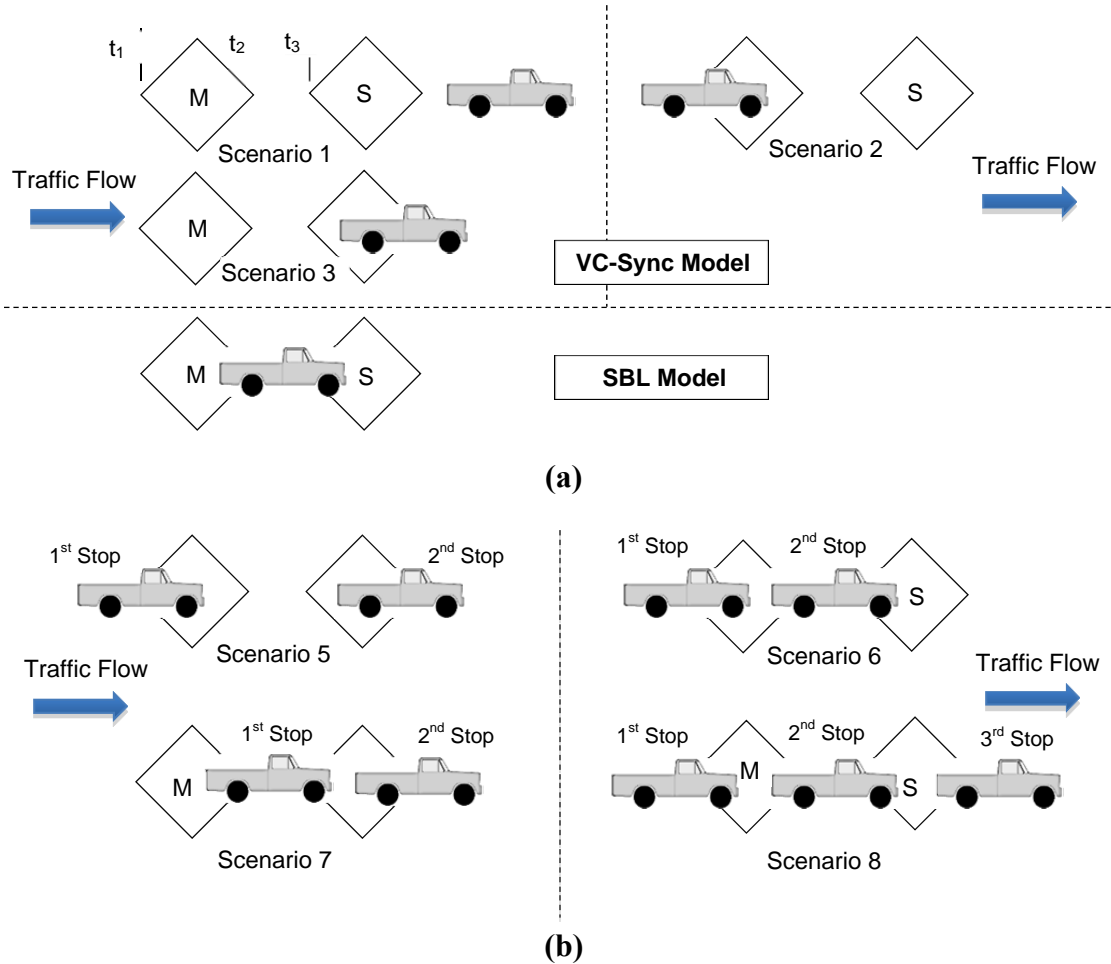


FIGURE 5 Scenarios of vehicle stopping on dual loops under congestion: (a) vehicle stop locations when one stop happened on loops and (b) vehicle stop locations when two or more stops happened on loops.

Statistical analysis of the sample data indicates that Scenarios 1 through 4 happened much more frequently than other scenarios (Figure 6 and Table 1). Scenarios 1 through 4 were hence focused in the study, and other scenarios will be considered in the future once sufficient sample data will be gained.

Under the stop-and-go traffic flow, a detected vehicle’s stopping status can be estimated based on its corresponding dual-loop data (i.e., the time stamps). An algorithm, as illustrated by Figure 7, was developed using On-times and difference of On-times to determine the scenario that the detected vehicle has fallen in. Based on the determined scenario, a suitable vehicle classification model can be applied to estimate the vehicle length.

In this algorithm, timestamp t_1, t_2, t_3, t_4 , OnT_1 , and OnT_2 are adopted as the variables. t_{s1} is defined as the threshold of OnT_1 and OnT_2 , and t_{s2} is defined as the threshold of the differences the timestamps. For a vehicle operating under the stop-and-go traffic condition:

1. If both of OnT_1 and OnT_2 are less than t_{s1} , it indicates that the vehicle did not make a stop within the dual-loop detection area, which means this vehicle falls into Scenario 1.
2. If OnT_1 is larger than t_{s1} , and OnT_2 is less than t_{s1} , it indicates that the vehicle spent much longer time on the upstream loop, and this vehicle will be identified into Scenario 2.
3. If OnT_1 is less than t_{s1} , and OnT_2 is larger than t_{s1} , it indicates that the vehicle spent much longer time on the downstream loop, and this vehicle will be identified into Scenario 3.
4. If both of OnT_1 and OnT_2 are larger than t_{s1} , and $t_3 - t_1 < t_{s2}$ and $t_4 - t_2 < t_{s2}$ (t_1, t_2, t_3 , and t_4 , are the same as defined previously), the vehicle can be identified as falling into Scenario 4.
5. In this study, in light of the statistical analysis on the dual-loop data under stop-and-go traffic, the thresholds are determined as: $t_{s1} = 4.1$ s, and $t_{s2} = 3.0$ s. A flow chart of the scenario identification algorithm is illustrated by Figure 7.

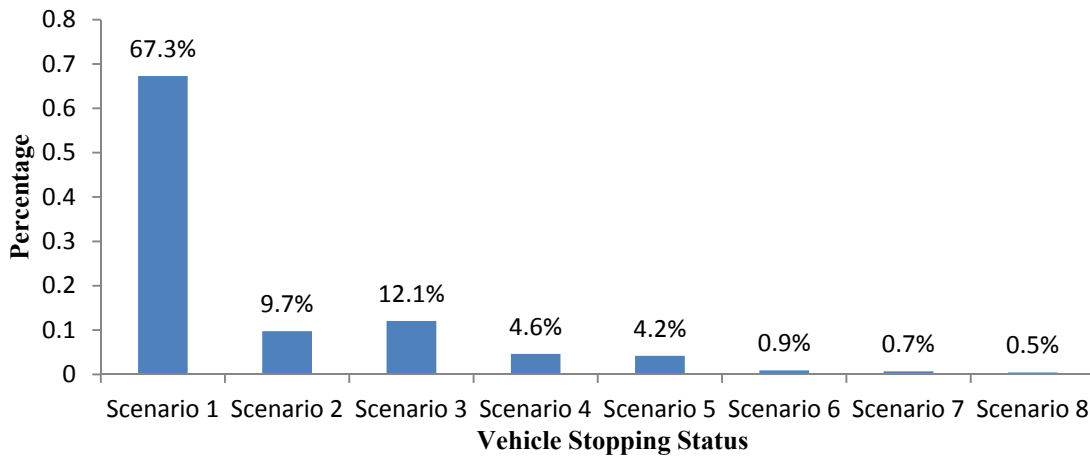


FIGURE 6 Percentage of vehicle stopping status in congested traffic.

TABLE 1 Vehicle Stopping Status Statistics

Scenario	1	2	3	4	5	6	7	8
Percentage	67.3%	9.7%	12.1%	4.6%	4.2%	0.9%	0.7%	0.5%

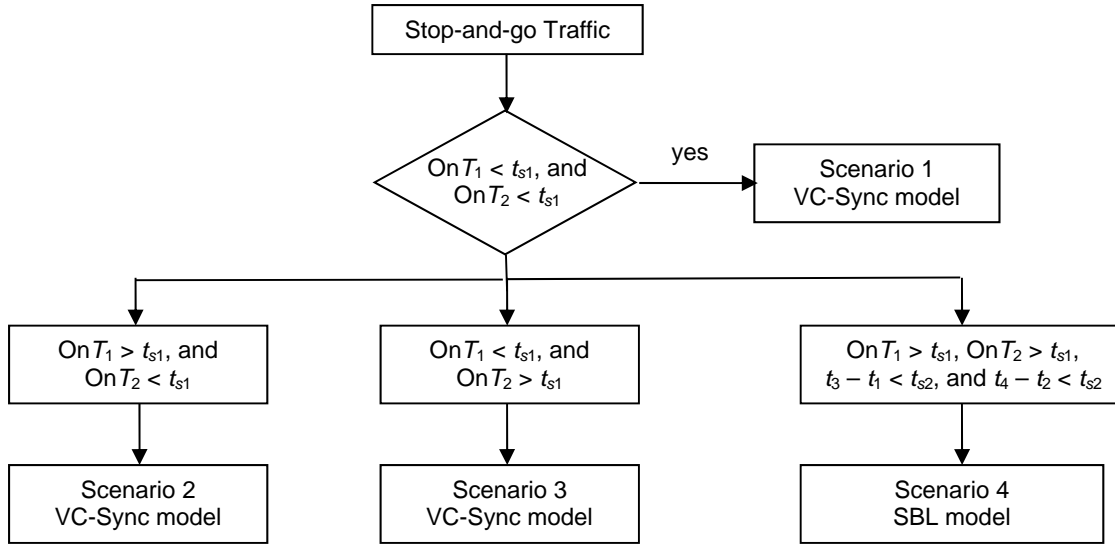


FIGURE 7 Scenario identification algorithm. (Note: t_{s1} is the threshold of OnT_1 and OnT_2 , and t_{s2} is the threshold of timestamp differences; $t_1, t_2, t_3, t_4, OnT_1$, and OnT_2 are the same as defined previously. In this study, t_{s1} and t_{s2} are determined as 4.1 and 3.0 s, respectively.)

LENGTH-BASED VEHICLE CLASSIFICATION MODELS UNDER CONGESTION

Vehicle Classification Model Under Synchronized Flow (Scenarios 1 Through 3)

Scenario 1 is a typical case of the synchronized traffic. Its flow density is higher than the free flow, and the freedom of maneuvers is greatly restricted. The travel speed is lower than the free flow, and higher than the stop-and-go flow. The VC-Sync model was proposed to estimate vehicle lengths under the synchronized traffic flow. In this model, a vehicle is assumed to pass the detection area at a constant acceleration rate a (a can be either positive or negative) without a stop. The length of the vehicle passing over the dual-loop detection area can be calculated by the equations as follows:

$$L_v = v_0 \cdot OnT_1 + \frac{1}{2} a (OnT_1)^2 - L_s \quad (5)$$

$$v_0 = \frac{D}{t} - \frac{a \cdot t}{2} \quad (6)$$

$$a = \frac{D}{t} \left[\frac{2 \cdot (OnT_1 - OnT_2)}{(OnT_2)^2 - (OnT_1)^2 + (OnT_1 + OnT_2) \cdot t} \right] \quad (7)$$

where

L_v = length of the detected vehicle (ft);

L_s = length of each single loop which makes up a dual-loop detector (ft);

v_0 = speed of the vehicle at the moment it is to enter the upstream loop (M loop) (ft/s);

a = vehicle acceleration (ft/s²); and
 D = distance between two loops (ft).

$t = t_3 - t_1$, $OnT_1 = t_2 - t_1$, and $OnT_2 = t_4 - t_3$, t_1 , t_2 , t_3 , and t_4 are time stamps when a vehicle enters or leaves the upstream loop (M loop) or downstream loop (S loop) (Figure 1).

Scenarios 2 and 3 can be viewed as special cases of Scenario 1. Scenario 2 is approximately equivalent to the situation in which a vehicle stops merely at the front edge of the upstream loop and then goes across the detection area without a further stop. This situation can be explained that a vehicle under the synchronized traffic is traversing through the detection area with acceleration and an initial speed of zero. Similarly, Scenario 3 is approximately equivalent to the situation in which a vehicle goes across the detection area without a stop and only stops at the end edge of the downstream loop. This situation can be interpreted that a vehicle under the synchronized traffic is traversing through the detection area with deceleration and a final speed of zero.

Vehicle Classification Model Under Stop-and-Go Flow (Scenario 4)

The stop-and-go traffic has much slower speeds, involving more frequent acceleration or deceleration maneuvers. Under the stop-and-go condition, a vehicle may stop within the detection area for at least once. Based on the ground-truth data, a statistical analysis was conducted to identify the pattern of vehicle stopping locations. As a result, the SBL model was developed to estimate the vehicle lengths under Scenario 4. For simplicity, it is assumed that the detected vehicle stops right in the middle of the dual loop. After stopping for a period of time t_s , the vehicle restarts to leave the dual-loop detection area at an acceleration rate a . The SBL model is expressed by Equation 8:

$$L_v = f_1 \cdot t_{dec} \cdot D \cdot \frac{1}{t} + \frac{1}{2} f_2 \cdot a \cdot t_{acc}^2 - L_s \quad (8)$$

where

$$\begin{aligned} t_{dec} + t_{acc} &= OnT_1 - t_s \text{ and } t_s = t_2 - t_3 - f_3 * t_{acc}^2 / v_{min}; \\ L_v &= \text{length of vehicle (ft);} \\ L_s &= \text{length of each single loop (ft);} \\ t_{dec} &= \text{time period as a vehicle enters the M loop until it stops (s);} \\ t_{acc} &= \text{time period as a vehicle starts to move and leaves the M loop (s);} \\ a &= \text{the average acceleration of vehicles as they start to move under} \\ &\quad \text{stop-and go traffic (ft/s}^2\text{);} \\ t_s &= \text{time period for a vehicle to stop on both loops (s);} \\ v_{min} &= \text{average minimum speed remaining without stop (ft/s);} \\ f_1, f_2, \text{ and } f_3 &= \text{adjusting factors for different vehicle types (in this study,} \\ &\quad f_1 = f_2 = f_3 = 1); \text{ and} \end{aligned}$$

D , t , t_2 , t_3 , OnT_1 , and OnT_2 = as the same as defined previously.

In order to make the SBL model applicable to estimating vehicle lengths in practice, the vehicle's acceleration rate (a) and average minimum non-stop speed (v_{min}) need to be

predetermined. In reality, however, it's extremely difficult to simply derive the acceleration rate of a detected vehicle from its corresponding dual-loop raw data under the stop-and-go condition. The GPS data collected by using GPS data loggers is therefore used to obtain a and v_{\min} . Based on the collected GPS data, the variables involved in the SBL model were eventually determined as follows: the average acceleration rate is 2.5 ft/s^2 and the average minimum speed v_{\min} is 7 ft/s .

Finally, the simulated vehicle lengths from the new developed models were compared with the results from the existing model while the ground-truth event data was used as a benchmark. The relative error is reduced from 33.5% of the existing model to 6.7% of the VC-Sync model under Scenarios 1 through 3 (see Figure 8). Under the stop-and-go traffic condition as represented by Scenario 4, the relative error was reduced from 235% of the existing model to 17.1% of the SBL model (Figure 9 and Table 2).

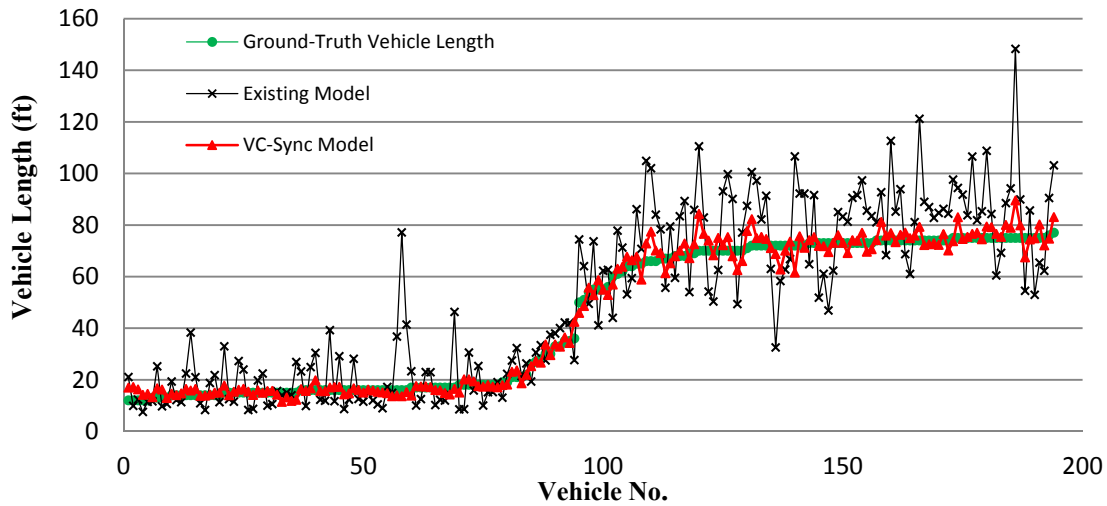


FIGURE 8 Estimated vehicle lengths under synchronized traffic.

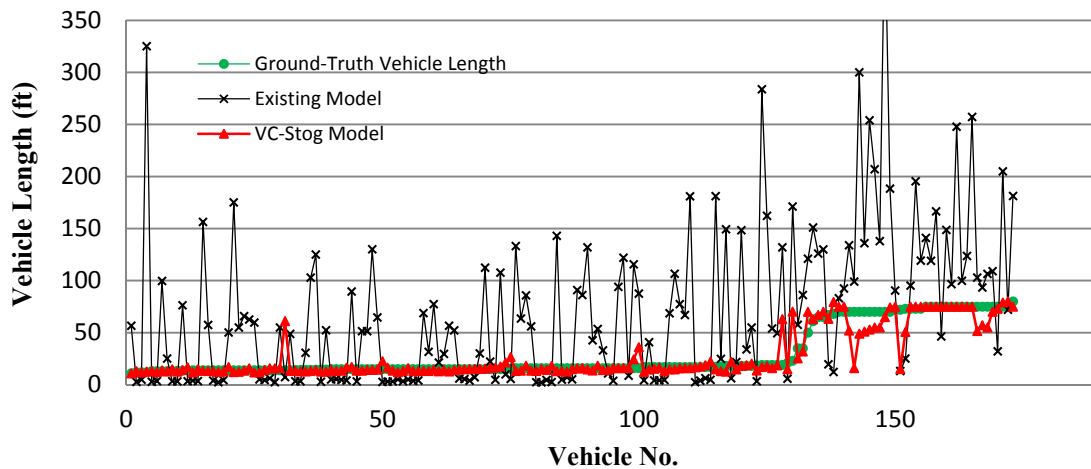


FIGURE 9 Estimated vehicle lengths under stop-and-go traffic.

TABLE 2 Relative Errors Produced by Classification Models

Traffic Flow Condition	Vehicle Classification Model	Error Produced
Synchronized flow	VC-Sync Model	6.7%
	Existing Model	33.5%
Stop-and-go flow	SBL Model	17.1%
	Existing Model	235%

CONCLUSION

The scenario-based vehicle classification models against both synchronized and stop-and-go traffic flows were developed by fully considering the impact of congested traffic flows. On the basis of watching synchronizing the ground-truth vehicular trajectory data and the dual-loop data, eight possible scenarios were synthesized based on possible stopping locations of the detected vehicles within the detection area. Those eight scenarios reflect the situations of vehicle stopping over loops, which were observed to occur with high possibility in the dual-loop detection area. This synthesized method simplifies the modeling of the vehicles' movements to reveal the impact of traffic on the identification of vehicle lengths at the dual-loop station. Under the synchronized traffic flow, acceleration or deceleration is considered in the VC-Sync model to reflect the speed variation between both loops, which were not conventionally considered in the existing applied models. As a result, the error of the vehicle length estimation is reduced from 33.5% by using the existing model to 6.7% by using the VC-Sync model. Under the stop-and-go traffic condition, the stopping status was synthesized into typical scenarios in the SBL model, which makes it easier to identify the variables involved in the associate vehicle length modeling. As a result, the error is reduced by using the SBL model from 235% to 17.1%, compared with the existing applied model.

Capability of identifying traffic phases is a critical support to applying the length-based vehicle classification models. This paper presents an innovative method for identifying the traffic phases that was developed based on integrated analysis of the existing traffic stream models and the new findings from the authors' empirical data analysis and modeling efforts. As a result, a heuristic traffic phase identification model has developed and successfully applied in the case study for evaluating the new length-based vehicle classification models with dual-loop data.

ACKNOWLEDGMENTS

The research presented in this paper was partially supported by a grant funded by Ohio Transportation Consortium. The authors thank Benjamin Coifman at Ohio State University for his assistance in providing the event dual-loop data.

REFERENCES

1. Nihan, N. L., X. Zhang, and Y. Wang. Evaluation of Dual-Loop Data Accuracy Using Video Ground Truth Data. Research Report WA-RD 535.1, Washington State Transportation Center, 2002.
2. Wei, H., E. Meyer, J. Lee, and C. E. Feng. Video-Capture-Based Approach to Extract Multiple

- Vehicular Trajectory Data for Traffic Modeling. *ASCE Journal of Transportation Engineering*, Vol. 131, No. 7, 2005, pp. 496–505.
3. Ai, Q. Length-Based Vehicle Classification Using Dual-loop Data under Congested Traffic Conditions. Ph.D. Dissertation. The University of Cincinnati. December 2013.
 4. Greenshields, B. D. A Study in Highway Capacity. *HRB Proc.*, Vol. 14, HRB, National Research Council, Washington, D.C., 1935, p. 468.
 5. Greenberg, H. An Analysis of Traffic Flow. *Operation Research*, Vol. 7, 1959, pp. 78–85.
 6. Underwood, R. T. Speed, Volume, and Density Relationships. In *Quality and Theory of Traffic Flow*. Yale Bureau of Highway Traffic, 1961, pp. 141–188.
 7. Edie, L. C. Car Following and Steady-State Theory for Non-congested Traffic. *Operations Research*, 9, 1961, pp. 66–76.
 8. Koshi, M., M. Iwasaki, and I. Okhura. Some Findings and an Overview on Vehicular Flow Characteristics. *Proc., 8th International Symposium on Transportation and Traffic Flow Theory*, University of Toronto Press, Toronto, Canada, 1983, pp. 403–426.
 9. May, A. D. Traffic Flow Fundamentals. Prentice-Hall. Englewood Cliffs, New Jersey. 1990.
 10. Hall, F. L. and M. A. Gunter. Further Analysis of the Flow-Concentration Relationship. In *Transportation Research Record 1091*, TRB, National Research Council, Washington, D.C., 1986, pp. 1–9.
 11. Polus, A., and M. Pollatschek. Stochastic Nature of Freeway Capacity and Its Estimation. *Canadian Journal of Civil Engineering*, 29, 2002, pp. 842–852.
 12. Kerner, B. S., and P. Konhäuser. Structure and Parameters of Clusters in Traffic Flow. *Physical Review E*, Vol. 50, No. 1, July 1994.
 13. Kerner, B. S., and S. L. Klenov. Explanation of Complex Dynamics of Congested Traffic in NGSIM-Data with Three-Phase Traffic Theory. Compendium of Papers CD-ROM, 89th Transportation Research Board Annual Meeting, Washington, D.C., January 2010.
 14. Bertini, B. A., and M. T. Leal. Empirical Study of Traffic Features at a Freeway Lane Drop. *Journal of Transportation Engineering*, 131(6), 2005, pp. 397–407.
 15. Zielke, B., R. Bertini, and M. Treiber. Empirical Measurement of Freeway Oscillation Characteristics: An International Comparison. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2088*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 57–67.
 16. Ahn, S., and M. J. Cassidy. Freeway Traffic Oscillation and Vehicle Lane-Change Maneuvers. *Transportation and Traffic Theory 2007*, Elsevier, 2007, pp. 691–710.
 17. Daganzo, C. F. A Behavioral Theory of Multi-lane Traffic Flow, Part I: Long Homogeneous Freeway Sections. *Transportation Research Part B*, 36(2), 2002, pp. 131–158.
 18. Mauch, M., and M. J. Cassidy. Freeway Traffic Oscillations: Observations and Predictions. *Transportation and Traffic Theory in the 21st Century: Proceedings of the Symposium of Traffic and Transportation Theory*. Elsevier, 2002, pp. 653–672.
 19. Treiber, M., and A. Kesting. Evidence of Convective Instability in Congested Traffic Flow: A Systematic Empirical and Theoretical Investigation. *Transportation Research Part B*, 45(9), 2011, pp. 1362–1377.
 20. Blandin, S., J Argote, A. Bayen, and D. Work. Phase Transition Model of Non-stationary Traffic Flow: Definition, Properties, and Solution Method. *Transportation Research Part B*, 52, 2013, pp. 31–55.
 21. Habib-Mattar, C., A. Polus, and M. A. Cohen. Analysis of the Breakdown Process on Congested Freeways. Compendium of Papers CD-ROM, 88th Transportation Research Board Annual Meeting, Washington, D.C., January 2009.
 22. Chow et al. 2010.
 23. Lorenz, R. M., and L. Elefteriadou. Defining Freeway Capacity as a Function of the Breakdown Probability. Compendium of Papers CD-ROM, 80th Transportation Research Board Annual Meeting, Washington, D.C., January 2001.

24. Elefteriadou, L., and P. Lertworawanich. Defining, Measuring, and Estimating Freeway Capacity. Compendium of Papers CD-ROM, 82nd Transportation Research Board Annual Meeting, Washington, D.C., January 2003.
25. Wieczorek, J., R. J. Fernández-Moctezuma, and R. L. Bertini. Techniques for Validating an Automatic Bottleneck Detection Tool Using Archived Freeway Sensor Data. Compendium of Papers CD-ROM, 89th Transportation Research Board Annual Meeting, Washington, D.C., January 2010.
26. Zhang, H. M., W. and Shen. A Numerical Investigation of the Stop-and-Go Traffic Patterns Upstream of a Freeway Lane-drop. Compendium of Papers CD-ROM, 88th Transportation Research Board Annual Meeting, Washington, D.C., January 2009.
27. Deng, W., H. Lei, and X. Zhou. Traffic State Estimation and Uncertainty Quantification Based on Heterogeneous Data Sources: A Three Detector Approach. *Transportation Research Part B*, 57, 2013, pp. 132–157.
28. Chen, L., and A. May. Traffic Detector Errors and Diagnostics. In *Transportation Research Record 1132*, TRB, National Research Council, Washington, D.C., 1987, pp. 82–93.
29. Turner, S., L. Albert, B. Gajewski, and W. Eisele. Archived Intelligent Transportation System Data Quality: Preliminary Analysis of San Antonio Transguide Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1719*, TRB, National Research Council, Washington, D.C., 2000, pp. 77–84.
30. Coifman, B. *Research Reports: An Assessment of Loop Detector and RTMS Performance*. Report No. UCB-ITS-PRR-2004-30, ISSN 1055-1425, 2004.
31. Cheevarunothai, P., Y. Wang, and L. N. Nihan. Development of Advanced Loop Event Analyzer (ALEDA) for Investigations of Dual-Loop Detector Malfunctions. Presented at the 12th World Congress on Intelligent Transportation Systems, San Francisco, 2005.
32. Nihan, N., X. Zhang, and Y. Yang. *Data: Quick Detection of Malfunctioning Loops and Calculation of Required Adjustments. Research Report*, TNW2006.
33. Fekpe, E., D. Gopalakrishna, and D. Middleton. *Highway Performance Monitoring System Traffic Data for High-Volume Routes: Best Practices and Guidelines*. Final Report. Office of Highway Policy Information Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., 2004.
34. Li, B. Bayesian Inference for Vehicle Speed and Vehicle Length Using Dual-Loop Detector Data. *Transportation Research Part B*, 44, 2010, pp. 108–119.
35. *Highway Capacity Manual 2010*. Transportation Research Board of the National Academies, Washington, D.C., 2010.

Additional Resources

- Athol, P. Interdependence of Certain Operational Characteristics within a Moving Traffic Stream. *Highway Research Record 72*, HRB, National Research Council, Washington, D.C., 1965, pp. 58–87.
- Cheevarunothai, P., and Y. Wang. Identification and Correction of Dual-Loop Sensitivity Problems. Compendium of Papers CD-ROM, 85th Transportation Research Board Annual Meeting, Washington, D.C., January 2006.
- Kerner, B. S. Experimental Features of Self-Organization in Traffic Flow. *Physical Review Letters*, Vol. 81, No.17, October 1998.
- Kerner, B. S. *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory*. Springer, 2004.
- Wei, H. *Characterize Dynamic Dilemma Zone and Minimize Its Effect at Signalized Intersections*. Ohio Transportation Consortium Research Project Report. 2008.

TRAFFIC FLOW MODELS

Optimal Velocity Model with Dual Boundary Optimal Velocity Function

HAO WANG

*Southeast University, China
University of California, Davis*

This paper proposes a dual boundary optimal velocity model (DBOVM) by substituting a dual boundary optimal velocity function (DBOVF) for the original one in the optimal velocity model (OVM). The proposed DBOVM can describe the driving behavior of accepting a range of satisfied conditions instead of an optimal one under steady traffic. The speed adjustment mechanism is introduced into the DBOVM, by which traffic flow can reach the steady state everywhere inside of the dual boundary steady region. Properties of traffic state transition in the law of DBOVM are analyzed. The approximately linear path of state transition is found, and four typical state transition patterns are presented. Besides, the stability of DBOVM is studied by means of numerical simulations. It is found that the dual boundary steady region has the hysteresis effect that is similar to the explicit time delay in the OVM. The speed adjustment mechanism can restrain the hysteresis to some extent and improve the traffic stability. The dual boundary steady region in the general DBOVM allows the traffic flow to reach some new steady state slightly apart from the formal one under the effect of small perturbation, which does not exist for models containing one-dimensional optimal velocity functions.

INTRODUCTION

Car-following models, as one of the most useful tools for describing traffic dynamics, have been developed for more than six decades. There are two main objectives in the car-following process: (1) reducing the speed difference and (2) maintaining an appropriate spacing between the following vehicle and the leading vehicle. Most early models as represented by Gazis-Herman-Rothery models were developed based on the first objective but failed to describe the second one. Newell (1) proposed a different model, which successfully captures the characteristic of car-following behaviors in maintaining an optimal spacing corresponding to the driving speed. However, because of the speed expression of the Newell model, it is not convenient to be used in traffic simulations. Thirty years later, OVM was developed (2, 3). Similar to the Newell model, the OVM contains the optimal velocity function, which allows the following vehicle to adjust its speed toward the optimal one, and consequently maintain the appropriate spacing. Moreover, the OVM does not have a time delay in its model expression, which makes it convenient for theoretical analysis. Because of these advantages, the OVM has drawn widespread attention during the last 20 years (4, 5).

The optimal velocity function assumes that there is a one-to-one correspondence between the spatial headway and the optimal driving speed in steady traffic state. However, such an assumption may be too ideal from the driver's perspective (6). Experience demonstrates that drivers are satisfied with a range of conditions instead of an accurate optimal performance.

Actually, this fact was also noticed in some early studies related to the psychophysical or action point models (7). The action point models classify the driving conditions into several scenarios separated by threshold boundaries in the “spacing–relative speed” diagram. Drivers will not adopt acceleration or deceleration until the changes in conditions exceed the thresholds of drivers’ perception. For instance, the Fritzsche model (8) and the Wiedemann model (9) have a two-dimensional zone in the spacing–relative speed diagram, within which drivers are satisfied with current conditions and maintain their speeds regardless of the performance of leading vehicles. Though the fundamentals of the action point model are closer to real driving behaviors, the complexity of model structure and the difficulty in calibration of thresholds restrict its application in theoretical research areas.

Another question about the assumption of optimal velocity function in the OVM came from the macroscopic observations of freeway traffic flow in recent years. Kerner and Rehborn (10, 11) first reported the widely scattered data of congested traffic in the fundamental diagram. Moreover, it was found that even the unstable data being removed, the remained steady traffic data still scatter in a two-dimensional area in the flow–density plot. These empirical findings indicate that there might be an acceptable range of spacing within which drivers are satisfied. In order to model such phenomenon, several microscopic traffic flow models were developed in recent years (12–15). Though solutions of steady state in these models occupy two-dimensional areas, the numbers of parameters are more than the OVM, which increases the complexity of models and reduces the efficiency.

In this paper, a DBOVM is proposed with the original optimal velocity function replaced by a DBOVF. The DBOVF is determined by two optimal velocity functions with different parameter values, which provides a range for spacing choosing in steady state. By introducing the DBOVF into the original OVM, the new model allows drivers to reach their steady states within a wide region instead of a specific optimal solution. Besides, the DBOVM also shows some interesting properties different from the original OVM, which are also worthy of attention.

The rest of the paper is organized as follows. First, the basic DBOVM and a simple example of DBOVF are presented in Section 2, and then stability analysis based on numerical simulation is conducted. In Section 3, a more general form of DBOVM is proposed, and some related model properties are analyzed. Some results of comparison analysis between the DBOVM and the original OVM are also provided through numerical simulations. Finally, the possibility of model extension is discussed.

BASIC DBOVM

The Model

Considering the facts that drivers would like to accept a range of spacing instead of an optimal one, it is assumed that the steady state occupies a two-dimensional area in the speed–spacing diagram. As shown in [Figure 1](#), there are two boundaries in the steady state region. Each boundary can be formulated by a certain type of optimal velocity function. The two boundaries of the steady state divide the speed–spacing diagram into three regions. In Region I, the spacing is too small for the driver to accept, and the driver will reduce the speed toward the optimal speed indicated by the left boundary optimal velocity function. In Region III the spacing is too

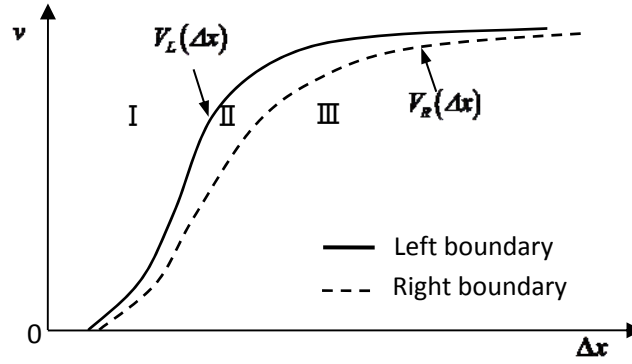


FIGURE 1 Illustration of DBOVF.

large, and the driver will accelerate toward the optimal speed indicated by the right boundary optimal velocity function. In Region II the driver is satisfied with current conditions and will not change the speed until the vehicle moves out of this steady region.

According to above assumptions, the basic DBOVM is expressed as follows.

$$\ddot{x}_n(t) = \begin{cases} \kappa \{V_L(\Delta x_n) - \dot{x}_n(t)\} & \text{if } : \dot{x}_n(t) > V_L(\Delta x_n) \\ 0 & \text{if } : V_R(\Delta x_n) \leq \dot{x}_n(t) \leq V_L(\Delta x_n) \\ \kappa \{V_R(\Delta x_n) - \dot{x}_n(t)\} & \text{if } : \dot{x}_n(t) < V_R(\Delta x_n) \end{cases} \quad (1)$$

where $x_n(t)$ denotes the position of the n th vehicle, κ is the sensitivity parameter, $V_L(\Delta x_n)$ and $V_R(\Delta x_n)$ denote the optimal velocity functions of left boundary and right boundary respectively. The space headway is noted as $\Delta x_n(t) = x_{n-1}(t) - x_n(t)$.

Simple Example of DBOVF

Many types of optimal velocity functions were proposed during the history of traffic flow studies. Among them there are three typical types that were used most widely by researchers, namely, the convex type represented by the exponential function (1), the piecewise linear function represented by triangle fundamental diagram model (16), and the S-shape function (2). In order to make comparison with the original OVM, Bando's S-shape function is used as the boundary function to build the DBOVF. Before modeling the DBOVF, two requirements are considered as follows:

1. The range of spacing in the steady state increases with the speed increasing; and
2. For a given speed v_e , the smallest and largest spacing in steady state are $\Delta x_L = V_L^{-1}(v_e)$ and $\Delta x_R = V_R^{-1}(v_e)$, then the derivatives of the two boundary-optimal-velocity-functions at Δx_L and Δx_R should satisfy the following inequation: $V'_L(\Delta x_L) \geq V'_R(\Delta x_R)$.

The first requirement comes from the studies on psychophysical car-following models (17, 18). These studies indicated that drivers perceive spacing changes through changes on a

visual angle subtended by the vehicle ahead. Under such a concept, the range of spacing in steady state may also be perceived through a change on a visual angle, which is supposed to be nearly stable for various driving speeds. As the visual angle approximates to the ratio of the vehicle width to the spacing and the spacing generally increases with the speed increasing, the range of spacing is positively correlated to the driving speed.

The second requirement is from the consideration that the deceleration is usually stronger than the acceleration at the margin of steady state. Suppose a vehicle moves a small distance δ away from the left boundary of steady state Δx_L , with the speed remaining as $v_e = V_L(\Delta x_L)$. Then, the deceleration a_{dec} according to the DBOVM is

$$a_{dec} = \kappa[V_L(\Delta x_L) - \delta] - v_e \quad (2)$$

Expanding Equation 2 at Δx_L and ignoring the higher order terms results in

$$a_{dec} = -\kappa\delta V'_L(\Delta x_L) \quad (3)$$

When the vehicle moves a small distance δ away from the right boundary of steady state Δx_R , the acceleration a_{acc} is

$$a_{acc} = \kappa\delta V'_R(\Delta x_R) \quad (4)$$

Therefore, the requirement of $V'_L(\Delta x_L) \geq V'_R(\Delta x_R)$ ensures the asymmetry between the acceleration and the deceleration, in other words, $|a_{dec}| \geq |a_{acc}|$.

The S-shape optimal velocity function used in Bando's work (3) is given as

$$V(\Delta x) = V_1 + V_2 \tanh[C_1 \Delta x - C_2] \quad (5)$$

where all parameters, V_1 , V_2 , C_1 , C_2 are positive values. Based on the requirements discussed above, a simple DBOVF follows:

$$V(\Delta x) = V_1 + V_2 \tanh[C_1 \Delta x - C_2], \quad C_1 \in [C_{1R}, C_{1L}] \quad (6)$$

and the two boundary-optimal-velocity-functions are given by

$$\text{Left boundary: } V_L(\Delta x) = V_1 + V_2 \tanh[C_{1L} \Delta x - C_2] \quad (6a)$$

$$\text{Right boundary: } V_R(\Delta x) = V_1 + V_2 \tanh[C_{1R} \Delta x - C_2] \quad (6b)$$

The parameter C_1 in the original OVM is replaced by C_{1L} and C_{1R} , which correspond to the left boundary and the right boundary, respectively. The proposed DBOVF has a very simple model structure and satisfies all the requirements listed above. In the following parts of this section, the authors will use it together with Equation 1 as the basic DBOVM to explore the properties of the model.

Stability Features of the Basic DBOVM

As a multiphase car-following model, the DBOVM does not have a uniform model expression, which makes it difficult for the analytical stability analysis. In view of this fact, numerical simulations are used to study the stability features of the basic DBOVM.

Local Stability

Treiber and Kesting (19) gave a detailed theoretical analysis on traffic stability in their recent book. It is pointed out that all time-continuous car-following models with a negative derivative of acceleration with respect to speed are unconditionally locally stable, if there are no explicit reaction times in models. Recall the dual-boundary-optimal-velocity-function displayed in Figure 1, the basic DBOVM satisfies the criterion suggested by Treiber and Kesting when the local traffic state is located outside of the dual boundary steady region. However, as the driver does not perform any acceleration within the steady region, it delays the driver's response to the leading vehicle when the traffic state moves through the dual boundary region in the speed-spacing phase diagram. Therefore, the basic DBOVM is analogous to the OVM with explicit delay (3) in some extent. The simulation studies on the local stability of the basic DBOVM are as follows.

In order to make a comparison study, we use the same parameters as the early literature (3) used for the original OVM, namely, $V_1 = 15.3$ m/s, $V_2 = 16.8$ m/s, $C_1 = 0.086$ m⁻¹, $C_2 = 2.1$, and $\kappa = 2.9$ s⁻¹. For the basic DBOVM, $C_{1L} = 0.088$ m⁻¹, $C_{1R} = 0.076$ m⁻¹, and keep all the other parameters the same as in the original OVM.

Three vehicles are considered in the local stability studies. All vehicles are in steady state at the beginning of the simulation. For the studies on the basic DBOVM, the initial state of vehicles should satisfy either the left or the right boundary optimal velocity function. Otherwise, the following vehicles may not respond to the perturbation from the leading vehicle according to the law of the basic DBOVM. Therefore, three simulations are conducted for three different scenarios respectively, which are (1) initial state on the right boundary of steady region, (2) initial state on the left boundary of steady region, and (3) initial state satisfying the optimal function in original OVM. All three simulations begin with the driving speed of 10 m/s for all vehicles. Such a speed ensures that the initial condition satisfies the string stability criterion of the original OVM (2). Then a small perturbation is added on the leading vehicle by giving its position an instantaneous change (either increasing by 1 m or reducing by 1m). The simulations are conducted with the time step of 0.1 s, and results of the simulations are illustrated in Figure 2.

Focus on the two subplots in the left side of Figure 2, which show the speed-headway phase diagram and speed time series of the simulation for the first scenario. Data of the first following vehicle and the second following vehicle are presented in red and blue, respectively. The numbers of seconds counted from the beginning of the simulation are marked beside the data points in the phase diagram. Two black curves represent the left boundary and the right boundary of DBOVF, respectively. At the beginning ($t = 0$), both two following vehicles stay on the right boundary of steady region. The perturbation is added onto the leading vehicle at the first second. It is apparent that the state points spiral counterclockwise toward the initial steady state around the left and right boundaries. After 30 s, the two following vehicles are still unable to return to the initial steady state. Though the perturbation is quite small (less than 5% of the initial

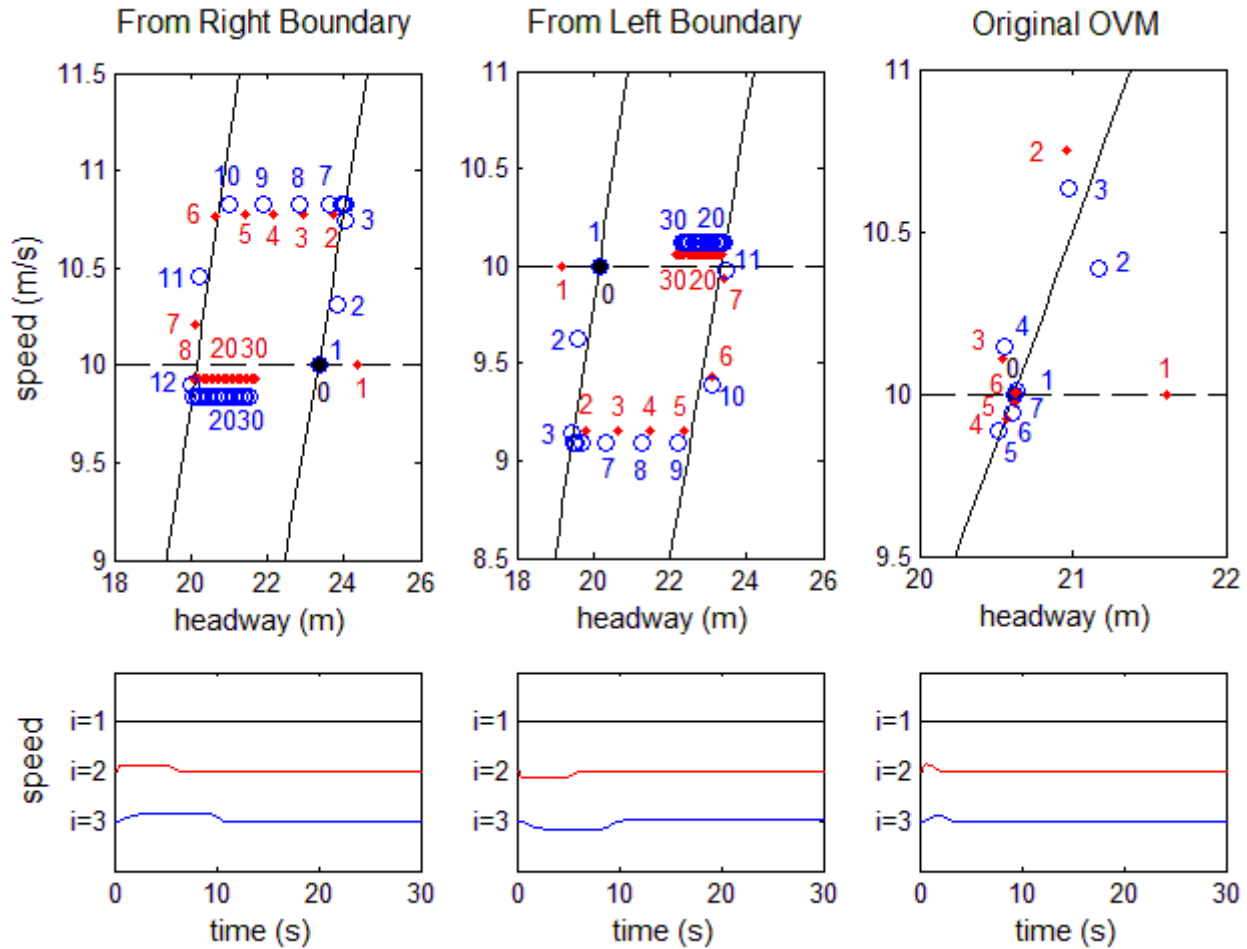


FIGURE 2 Phase diagram and speed time series of basic DBOVM.

spacing), it is slightly amplified during it propagating upstream. In the phase diagram the blue state transition trajectory of the second vehicle is outside of the red trajectory of the first vehicle.

In the second scenario, both two following vehicles stay on the left boundary of the steady region at the beginning. Similarly, the small perturbation is added onto the leading vehicle at the first second. The results of the simulation are displayed in the two subplots in the middle of Figure 2. There are similar dynamical properties as shown in the first scenario. The states of the following vehicles move clockwise toward the initial steady state in a very slow manner. Moreover, the amplification of perturbation is also observed during it propagates upstream.

The two subplots in the right-hand side of Figure 2 give the results of the simulation for the third scenario. Under the law of the original OVM, the following vehicles return to the initial steady state much faster through smaller spiral trajectories. Both two following vehicles are found to recover to their initial steady state 6 s after the perturbation is added.

Since the driver does not perform any acceleration in the two-dimensional region of steady state according to the basic DBOVM, it takes a long time to converge to the steady state. Such property may not have remarkable influence on the local stability of the basic DBOVM, whereas the slow convergence is likely to amplify the perturbation through the propagation. The following gives the simulation result of the dynamics of the basic DBOVM in a long platoon.

A Case of Perturbation Propagating Along Platoon

All parameters remained the same as in the local stability studies. The platoon on the open boundary road involves 200 vehicles in the simulation. The initial speed and spacing are 5 m/s and 18.2 m for all vehicles in the platoon, which satisfies the right boundary optimal velocity function of the two-dimensional steady region. Then the perturbation is added 10 s after the beginning of the simulation by giving the first vehicle's position a sudden increment of 1 m. The time step is 0.1 s. The initial conditions above ensure the string stability for both two-boundary optimal velocity functions under the law of the original OVM [i.e., $V'(\Delta x) \leq \kappa/2$]. However, simulation for the basic DBOVM shows that the platoon is string unstable, as illustrated in [Figure 3](#).

A very slight acceleration wave can be found propagating upstream after the perturbation occurs. Then a deceleration wave appears followed by another acceleration wave. After about 360 s from the beginning, the deceleration wave merges into the following acceleration wave, which is again followed by a new deceleration wave. From Figure 3 it is apparent that the small perturbation finally induces a stop-and-go wave during its propagation to the upstream of the platoon.

The simulation above indicates that the string stable criterion of the original OVM for the condition on boundary optimal velocity functions may not ensure the string stable of the basic DBOVM. In other words, the two-dimensional steady region in the basic DBOVM reduces the stability of traffic flow. The mechanism and some amendments about the model are given in the following parts of the paper.

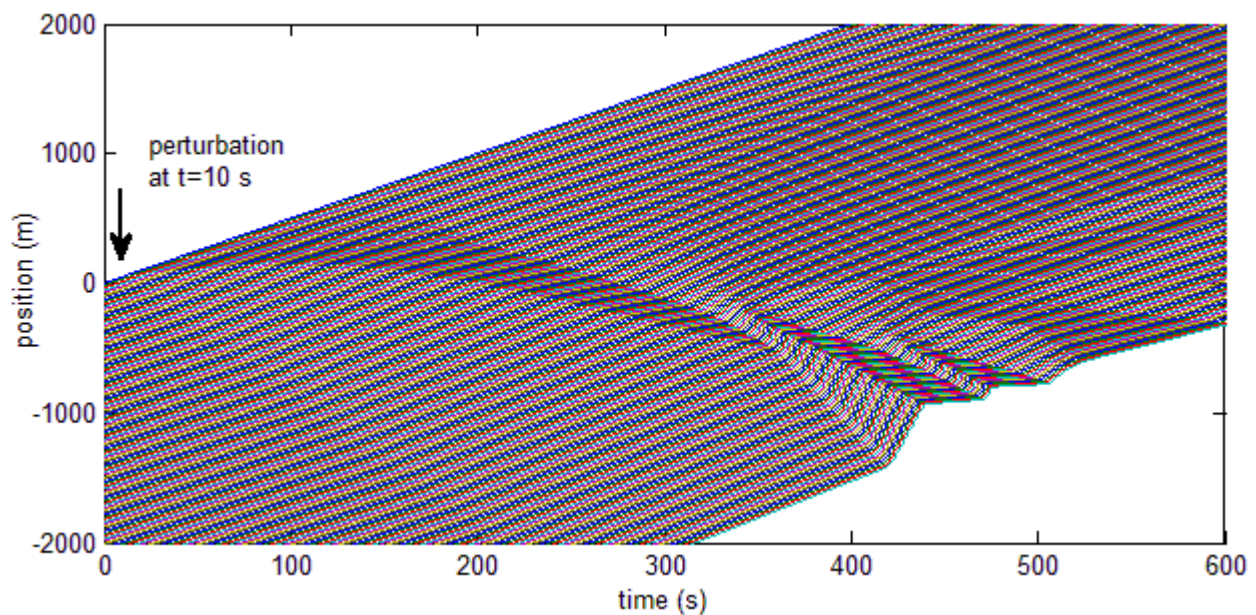


FIGURE 3 Propagation of small perturbation along the platoon in the basic DBOVM.

GENERAL DBOVM

The Model

In order to overcome the flaw in the basic DBOVM, the speed adjustment mechanism was introduced. In the new model, drivers are allowed to adjust their driving speeds toward the speed of leading vehicles within the dual boundary steady region. The modified model is given below.

$$\ddot{x}_n(t) = \begin{cases} \kappa \{V_L(\Delta x_n) - \dot{x}_n(t)\} & \text{if } : \dot{x}_n(t) > V_L(\Delta x_n) \\ \lambda \{\dot{x}_{n-1}(t) - \dot{x}_n(t)\} & \text{if } : V_R(\Delta x_n) \leq \dot{x}_n(t) \leq V_L(\Delta x_n) \\ \kappa \{V_R(\Delta x_n) - \dot{x}_n(t)\} & \text{if } : \dot{x}_n(t) < V_R(\Delta x_n) \end{cases} \quad (7)$$

where λ is the sensitivity parameter of the speed difference between the leader and the follower. The speed adjustment mechanism captures the fact that drivers try to duplicate the speed of the vehicle ahead and maintain a stable spacing once they drive into their satisfied range of conditions, namely, the dual boundary steady region. As the basic DBOVM can be regarded as a special case of the new model ($\lambda = 0$), the new model is called the general DBOVM.

Properties of State Transition

According to the general DBOVM, the dynamic properties of vehicles are just the same as under the law of the original OVM, when the traffic state locates outside of the dual boundary steady region in the speed-spacing diagram. In the following, the properties of state transition induced by the speed adjustment mechanism inside of the dual boundary region are analyzed.

Path of State Transition

Suppose the state of the n th vehicle is located inside of the dual boundary steady region at time t as shown in [Figure 4a](#). The speeds of the n th vehicle and the vehicle ahead are $\dot{x}_n(t)$ and $\dot{x}_{n-1}(t)$, respectively. According to the law of general DBOVM, the acceleration of the n th vehicle at time t is given by

$$\ddot{x}_n(t) = \lambda \{\dot{x}_{n-1}(t) - \dot{x}_n(t)\} \quad (8)$$

Assume that the vehicle moves with constant acceleration during a very small time interval τ , then speeds of the n th vehicle and $(n-1)$ th vehicle at time $t + \tau$ can be expressed as

$$\dot{x}_n(t + \tau) = \dot{x}_n(t) + \tau \ddot{x}_n(t) \quad (9a)$$

$$\dot{x}_{n-1}(t + \tau) = \dot{x}_{n-1}(t) + \tau \ddot{x}_{n-1}(t) \quad (9b)$$

Then, the change of the spacing between the n th vehicle and $(n-1)$ th vehicle during the

time interval τ can be calculated as follows:

$$\Delta x_n(t + \tau) - \Delta x_n(t) = \tau \{ \dot{x}_{n-1}(t) - \dot{x}_n(t) \} + \frac{\tau^2}{2} \{ \ddot{x}_{n-1}(t) - \ddot{x}_n(t) \} \quad (10)$$

Therefore, the slope of the state transition path from time t to $t + \tau$ in the speed–spacing phase diagram can be derived as

$$\beta = \frac{\dot{x}_n(t + \tau) - \dot{x}_n(t)}{\Delta x_n(t + \tau) - \Delta x_n(t)} \quad (11)$$

By substituting Equations 8, 9a, 9b, and 10 into Equation 11, one gets the following

$$\beta = \frac{1}{\frac{1}{\lambda} + \frac{\tau}{2} \left(\frac{\ddot{x}_{n-1}(t)}{\dot{x}_n(t)} - 1 \right)} \quad (12)$$

Equation 12 indicates that the slope β of the state transition path approximates to the sensitivity parameter λ in the general DBOVM when (1) the time step τ in the numerical calculation approaches zero, or (2) the following vehicle and the leading vehicle have similar accelerations. Specifically, when the speed of the leading vehicle remains stable [i.e., $\ddot{x}_{n-1}(t) = 0$], the slope of the state transition path is expressed as

$$\beta = \frac{1}{\frac{1}{\lambda} - \frac{\tau}{2}} \quad (13)$$

Considering that the time step used in numerical simulation for car-following models is usually quite small (e.g., $\tau = 0.1$ s), the slope of the state transition path can be approximately expressed as λ everywhere within the dual boundary steady region. As shown in Figure 4b, the dual boundary steady region involves numerous parallel linear paths of state transition, as represented by the red lines in the figure. Once the traffic states reach the boundaries of the steady region, the traffic states move along the state transition paths until converging at steady states. Figure 5 presents some simple scenarios of traffic states transitions in the speed-spacing phase diagram.

Suppose the leading vehicle's speed remains stable at $\dot{x}_{n-1}(t)$ during the traffic state transition of the following vehicle. The numbers in the figure denote the traffic states in terms of time series. The black solid dots represent the final steady states, which are always located at the horizontal line with the speed equal to $\dot{x}_{n-1}(t)$. Figure 5a, b, and c display the paths of state transitions from the start points (remarked by the number 1) to the final steady state (black solid dots). Figure 5a shows that the traffic state can converge to the steady state directly, if the initial state-transition-path along which the traffic state enters the dual boundary region has an intersection point with the speed horizontal line of the leading vehicle. Otherwise, the traffic state has to firstly move across the dual boundary region along the state-transition-path, and then

spirals into the dual boundary region again to reach the final steady point (see Figure 5b and 5c).

For most cases, there is at least one critical point on the boundary where the derivative is equal to the slope of the state transition path. Figure 5d illustrates a special case, in which four critical points (points A, B, C, and D) exist on the two boundaries as represented by the blue solid dots. Take the critical point A for example. When the speed of leading vehicle is lower than that of the speed at point A, all the initial states that are located in the right-hand side of point A cannot find paths within the dual boundary region for state transitions. Instead, the state transition paths for such initial states go zigzag against the boundary until passing the critical point A and then approach the final steady state along the state transition path. For the neighborhoods around other critical points, similar properties of the state transitions can be found, but it is not necessary to go into those details here.

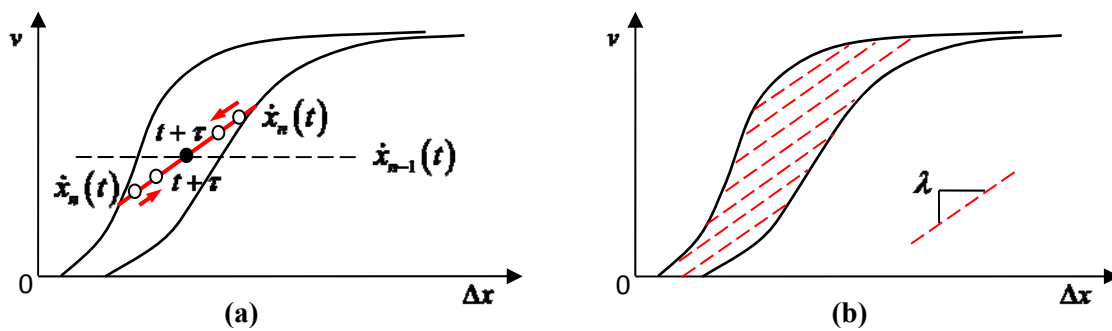


FIGURE 4 Paths of state transition within dual boundary steady region.

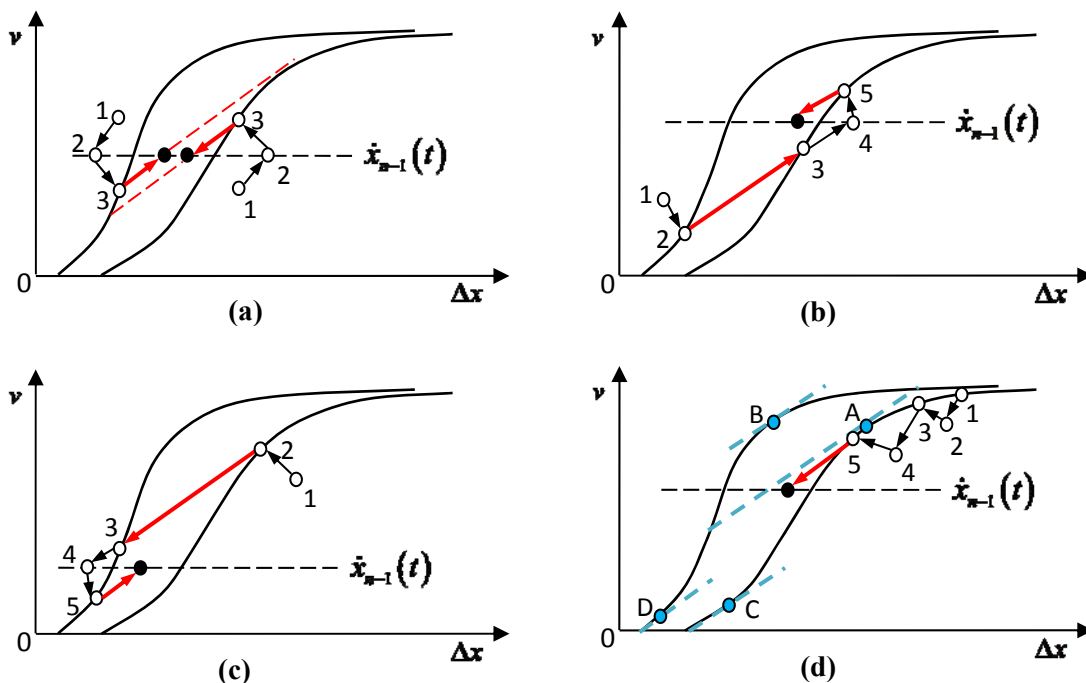


FIGURE 5 State transitions in speed-spacing phase diagram.

Convergence of Traffic State

After obtaining the state transition paths within the dual boundary steady region, another interesting question is in what manner the traffic state moves along the state transition path. Suppose the n th vehicle reaches the right boundary of the two-dimensional region at time t_0 in a deceleration process. The speeds of the n th vehicle and the vehicle ahead are denoted as $\dot{x}_n(t_0)$ and $\dot{x}_{n-1}(t_0)$, and thus $\dot{x}_n(t_0) > \dot{x}_{n-1}(t_0)$. For the sake of convenience, it is assumed that the leading vehicle keeps the speed $\dot{x}_{n-1}(t_0)$ stable. For a given time step τ , the speed of the n th vehicle at time $t_0 + \tau$ can be derived according to model Equation 7 as follows:

$$\dot{x}_n(t_0 + \tau) = \dot{x}_n(t_0) - \lambda\tau\{\dot{x}_n(t_0) - \dot{x}_{n-1}(t_0)\} \quad (14)$$

Similarly, the speed of the n th vehicle at time $t_0 + m\tau$ is

$$\dot{x}_n(t_0 + m\tau) = \dot{x}_n(t_0 + (m-1)\tau) - \lambda\tau\{\dot{x}_n(t_0 + (m-1)\tau) - \dot{x}_{n-1}(t_0 + (m-1)\tau)\} \quad (15)$$

By recursive method, the solution to Equation 15 is

$$\dot{x}_n(t_0 + m\tau) = \dot{x}_n(t_0)(1 - \lambda\tau)^m - \dot{x}_{n-1}(t_0)\{(1 - \lambda\tau)^m - 1\} \quad (16)$$

Then, the speed difference between the n th vehicle and the vehicle ahead at time $t_0 + m\tau$ can be expressed as

$$\Delta\dot{x}_n(t_0 + m\tau) = \dot{x}_{n-1}(t_0) - \dot{x}_n(t_0 + m\tau) = \{\dot{x}_{n-1}(t_0) - \dot{x}_n(t_0)\}(1 - \lambda\tau)^m = \Delta\dot{x}_n(t_0)(1 - \lambda\tau)^m \quad (17)$$

Equation 17 indicates that the speed difference between the following vehicle and the leading vehicle drops by $(1 - \lambda\tau)$ in every time step. Specifically, the speed difference drops by $(1 - \lambda)$ per second, when the time step τ equal to 1 s. If $\lambda = 0.5 \text{ s}^{-1}$, the speed difference drops by half per second, which also means that the distance between the current state and the final steady state along the state transition path is shortened by half every second as shown in [Figure 6](#). Then, after 3 s, the speed difference drops to 1/8 times of the initial value, and the following vehicle is approaching stable.

Recalling the basic DBOVM proposed in Section 2, as the slope of the state transition path $\lambda = 0$, the speed difference cannot get reduced during the state transition within the dual boundary steady region. That's why the basic DBOVM always takes an unreasonably long time to approach the steady state. Moreover, the state transition paths in the basic DBOVM are horizontally distributed within the steady region, which indicates that the traffic state cannot reach the steady state theoretically. However, under the speed adjustment mechanism in general DBOVM, the traffic state can converge to any spacing in between the dual boundary steady region.

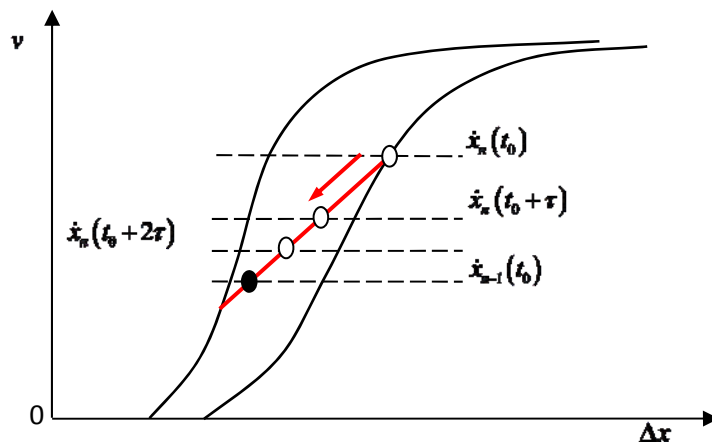


FIGURE 6 Process of state transition with $\lambda = 0.5$.

Stability Features of the General DBOVM

Local Stability

Using the same conditions as in the studies on the basic DBOVM, the sensitivity parameter λ is set as $\lambda = 0.5 \text{ s}^{-1}$. Similar to the studies in Section 2, two simulations are conducted, one for the initial state on right boundary and the other for left boundary. Figure 7 shows the simulation results of both scenarios. The red data and the blue data represent the first following vehicle and the second following vehicle, respectively, and the numbers denote the time series of the state points. It is found that the slopes of all state transition paths are equal to 0.51, which is consistent with the result computed by Equation 13. As the time step of 0.1 s is used in the simulation, the rate of speed difference dropping per second can be computed by Equation 17 as $(1 - 0.5 \times 0.1)^{10} = 0.60$. Therefore, in every second the speed difference drops to 0.60 times of it 1 s before, which is also consistent with the simulation results. After 6 s from the perturbation, the state of the first following vehicle is very close to the steady state. At the tenth second, the state of vehicle converges at some spacing between the two boundaries. Besides, it can be found that the final steady spacing of the second following vehicle is closer to the boundary in both scenarios, which proves that the effects of speed adjustment mechanism can allow the vehicle to reach any steady state within the dual boundary steady region.

The general DBOVM shows a similarity in the convergence speed with that of the original OVM, if one chooses an appropriate sensitivity parameter λ . In order to have deep insight of the dynamic features of the general DBOVM, numerical simulations on string stability studies are conducted in following.

String Stability

The periodical ring road is used as the simulation condition for studies on string stability. The platoon in the ring road involves 100 vehicles in total. Both the left boundary and the right boundary are considered as the initial steady state of the platoon in the ring. A small perturbation is added on the first vehicle by an instantaneous small jump (less than 5%) in position, and the

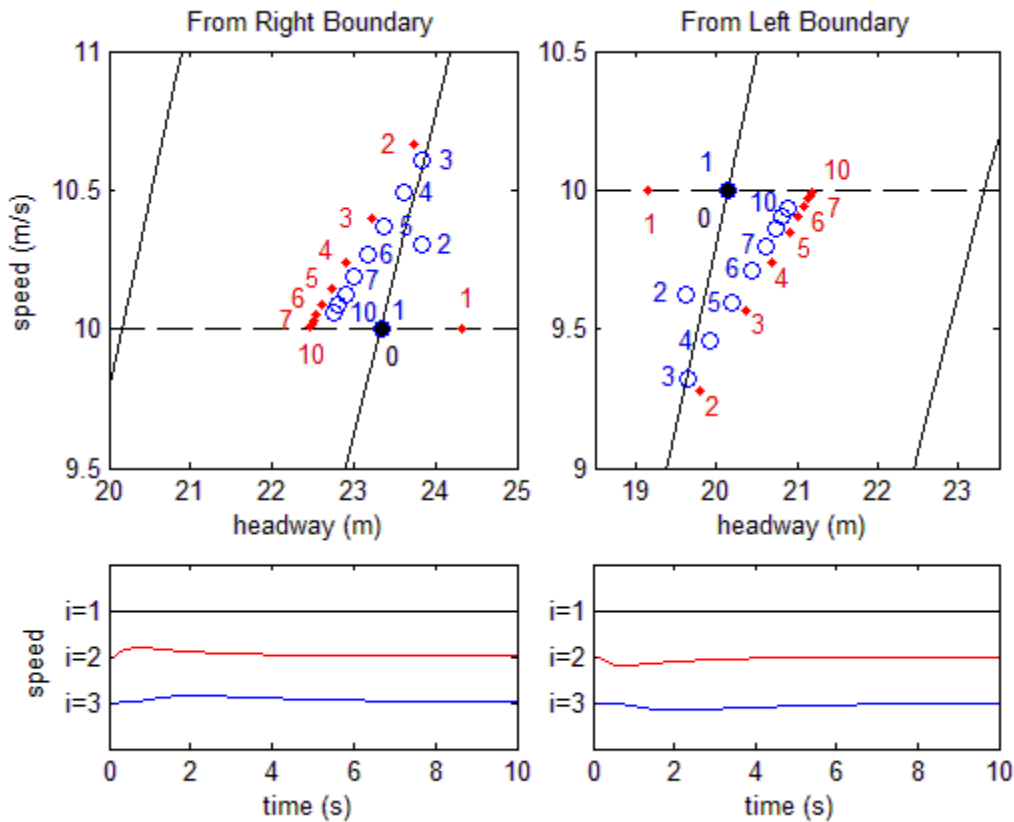


FIGURE 7 Simulations on local stability of the general DBOVM.

time step is 0.1 s in all simulations. In order to make comparison, simulations are conducted for all three models: the original OVM, the basic DBOVM ($\lambda = 0$), and the general DBOVM ($\lambda = 0.5$). All model parameters remained the same as in the local stability studies.

First the dynamics of the platoon under the unstable conditions are investigated. The initial speed of the platoon is 16 m/s, which is string unstable for both the left boundary and the right boundary under the law of OVM. Figure 8 gives the snapshots of the platoon's dynamics at the time $t = 5,000$ s from the time $t = 0$ at which the perturbation added. From the figure we can see it is apparent that all three models produce the stop-and-go patterns after 5,000 s. The hysteresis loop produced by the basic DBOVM is the largest in the phase diagram. With the effect of speed adjustment mechanism, the hysteresis loop of the general DBOVM is smaller, whereas the original OVM has the smallest one. Another important feature is the relation of the hysteresis loops between the original OVM and the basic DBOVM. In the simulation beginning from the left boundary of steady region, the left side of the hysteresis loop produced by the OVM overlaps the same side of the DBOVM's loop, while they overlap in the right side for the case from the right boundary.

The simulation results indicate that the dual boundary steady region in the DBOVM has a similar effect as the explicit delay for the original OVM (3), which can enlarge the hysteresis loop in traffic dynamics. However, the speed adjustment mechanism in the general DBOVM restrains such effect in some extent.

In the next step, the ring road simulation for every different initial steady state on the dual

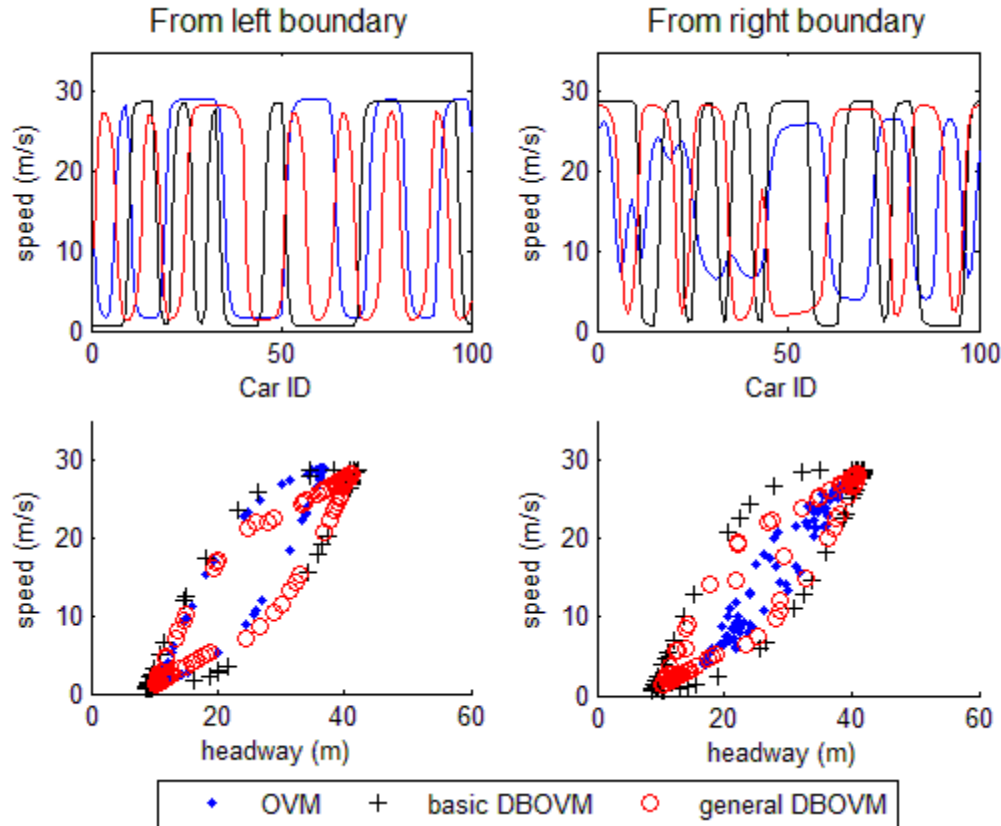


FIGURE 8 Snapshots of phase diagrams of three models at $t = 5,000$ s.

boundary (by the interval of 0.1 m/s in the initial speed) is repeated, and the total simulation time to 100,000 s is added, in order to ensure that the perturbation has sufficient time for evaluation. Table 1 illustrates the resulting string unstable region for the general DBOVM and the basic DBOVM.

The string unstable regions in the basic DBOVM ($\lambda = 0$) are wider than the original OVM because of the hysteresis effect induced by the dual boundary region. However, the string unstable regions for general DBOVM ($\lambda = 0.5$) are close to that of the original OVM, under the effects of speed adjustment mechanism. Because only small perturbations are considered in the simulations, the states of vehicles always move around the boundary where they get started, as illustrated in Figure 5a. Therefore, the string unstable regions of speed are not the same on the two boundaries but are related to the mathematical properties of the boundary optimal velocity function. Moreover, as the derivative of left boundary optimal velocity function is larger than that of the right boundary at the same speed, simulations starting from the left boundary are likely to produce stronger acceleration (or deceleration) than the cases beginning from the right boundary, which makes the left boundary less stable than the right one. Therefore, the unstable region of DBOVM on the left boundary is wider than that in the right boundary.

Another interesting phenomenon that can be found is that the final steady states that the platoon reaches after the perturbation are slightly different from the original ones where the platoon gets started. Specifically, the speed of the final steady state is higher than the initial

TABLE 1 String Unstable Regions in General DBOVM and Basic DBOVM

	Left Boundary OVF $V(\Delta x) = 15.3 + 16.8 \tanh [0.088 \Delta x - 2.1]$	Right Boundary OVF $V(\Delta x) = 15.3 + 16.8 \tanh [0.076 \Delta x - 2.1]$
DBOVM $\lambda = 0.5$	$\Delta x \in (17.9, 31.7)$	$\Delta x \in (21.1, 31.8)$
	$v \in (7.2, 25.3)$	$v \in (7.6, 20.5)$
DBOVM $\lambda = 0$	$\Delta x \in (14.3, 34.9)$	$\Delta x \in (15.4, 40.3)$
	$v \in (3.8, 27.9)$	$v \in (3.0, 27.8)$
OVM	$\Delta x \in (16.5, 31.2)$	$\Delta x \in (22.0, 35.2)$
	$v \in (5.7, 24.9)$	$v \in (8.5, 24.0)$

steady state in the order of 0.1 m/s, when the platoon starts from the right boundary. For the cases of starting from the left boundary, the final steady speed is found lower than the initial one also in the order of 0.1 m/s. The two-dimensional-steady-region allows the platoon to transfer its steady state under the perturbation. Similar shifts can be observed on the steady spacing of the platoon, if the simulation is conducted in an open boundary road instead of the ring road.

DISCUSSION

The basic DBOVM substitute the two-dimensional-steady-region for the optimal velocity function in OVM, which is reasonable for modeling the satisfaction range of steady state in car following. However, the basic DBOVM cannot reach the steady state inside of the dual boundary region during the dynamic process of traffic flow. Therefore, the amendment of the speed adjustment mechanism is necessary in the general DBOVM. The effect of speed adjustment in the proposed model is similar to the speed adaption concept in the literature (12), while the latter one is more complex in model structure. Besides, the general DBOVM has some similarities with the FVDM (5), which also has a speed difference term in the model. The speed adjustment effect in the general DBOVM only works within the dual boundary steady region, whereas the FVDM does not contain a two-dimensional-steady-region, and the speed difference term always contributes to the acceleration.

In this paper, only the S-shape function is considered for the expressions of the DBOVF. The differences in the functions of DBOVF may change the stability conditions of the general DBOVM, whereas the properties of traffic state transitions are the same as discussed in this paper. Another point that should be mentioned is the paths of state transitions within the dual boundary steady region are not strictly but approximately linear. However, one can still use the sensitivity parameter λ as the slope of the state transition path for roughly analysis.

SUMMARY

The main contribution of this paper is that the authors proposed a simple car-following model called general DBOVM, which can describe the driving behavior of accepting a range of satisfied conditions instead of an optimal one under steady traffic. The model is developed based on the OVM, with only two additional parameters. Therefore, it is very convenient for analytical and numerical analyses.

A simple speed adjustment mechanism is introduced into the basic DBOVM, with which the traffic state can converge to steady state everywhere inside of the two boundary steady region. Under the effect of speed adjustment, traffic states are transferred along some specific paths with an approximately constant slope equal to the sensitivity parameter of the speed difference term in general DBOVM.

The dual boundary steady region in DBOVM has the hysteresis effect, which is similar to the effect of explicit delay in OVM. The wider the dual boundary steady region is, the stronger the hysteresis effect will be. In spite of the instability that results from the hysteresis of dual boundary region, the speed adjustment effect in general DBOVM restrains the hysteresis and improves the stability of traffic.

The dual boundary steady region in the general DBOVM allows the traffic flow to reach a new steady state slightly apart from the formal one under the effect of small perturbation. This property does not exist in models with one dimensional optimal velocity functions.

The authors present only the framework of general DBOVM, and such a dual boundary steady region can also be introduced into other well-known car-following models. Moreover, the parameters of the general DBOVM are required to be calibrated by real traffic flow data, and applications of the proposed model will be studied in the next step.

ACKNOWLEDGMENTS

This work is supported by National Natural Science Foundation of China. Thanks also to reviewers for helping improve this paper.

REFERENCES

1. Newell, G. F. Nonlinear Effects in the Dynamics of Car Following. *Operation Research*, Vol. 9, No. 2, 1961, pp. 209–229.
2. Bando, M., K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, A Dynamical Model of Traffic Congestion and Numerical Simulation. *Physical Review E*, Vol. 51, No. 2, 1995, pp. 1035–1042.
3. Bando, M., K. Hasebe, K. Nakanishi, and A. Nakayama. Analysis of Optimal Velocity Model with Explicit Delay. *Physical Review E*, Vol. 58, 1998, pp. 5429–5435.
4. Helbing, D., and B. Tilch. Generalized Force Model of Traffic Dynamics. *Physical Review E*, Vol. 58, No. 1, 1998, pp. 133–138.
5. Jiang, R., Q.-S. Wu, and Z. J. Zhu. Full Velocity Difference Model for Car-Following Theory. *Physical Review E*, Vol. 64, 2001, pp. 017101.
6. Boer, E. R. Car Following from the Driver's Perspective. *Transportation Research F*, Vol. 2, 1999, pp. 201–206.
7. Brackstone, M., and M. McDonald. Car-following: A Historical Review. *Transportation Research F*, Vol. 2, 1999, pp. 181–196.
8. Fritzsche, H. T. A Model for Traffic Simulation. *Traffic Engineering and Control*, Vol. 35, No. 5, 1994, pp. 317–321.
9. Wiedemann, R., and U. Reiter. *Microscopic Traffic Simulation: The Simulation System MISSION, Background and Actual State*. CEC Project ICARUS (V1052), Final Report, Vol. 2, Appendix A. Brussels: CEC, 1992.
10. Kerner, B. S., and H. Rehborn. Experimental Features and Characteristics of Traffic Jams. *Physical Review E*, Vol. 53, 1996, pp. 1297–1300.

11. Kerner, B. S., and H. Rehborn. Experimental Properties of Complexity in Traffic Flow. *Physical Review E*, Vol. 53, 1996, pp. 4275–4278.
12. Kerner, B. S., and S. L. Klenov. A Microscopic Model for Phase Transitions in Traffic Flow. *Journal of Physics A*, Vol. 35, 2002, pp. 31–43.
13. Kerner, B. S., and S. L. Klenov. Deterministic Microscopic Three-phase Traffic Flow Models. *Journal of Physics A*, Vol. 39, 2006, pp. 1775–1809.
14. Davis, L. C. Multilane Simulations of Traffic Phases. *Physical Review E*, Vol. 69, 2004, id. 016108.
15. Gao, K., R. Jiang, B. H. Wang, and Q.-S. Wu. Discontinuous Transition from Free Flow to Synchronized Flow Induced by Short-range Intersection between Vehicles in a Three-Phase Traffic Flow Model. *Physica A*, Vol. 388, 2009, pp. 3233–3243.
16. Daganzo, C. F. The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research B*, Vol. 28, No. 4, 1994, pp. 269–287.
17. Michaels, R. M. Perceptual Factors in Car Following. In *Proceedings of International Symposium on the Theory of Road Traffic Flow*, 1963, pp. 44–59.
18. Evans, L., and R. Rothery. Perceptual Thresholds in Car Following: A Recent Comparison. *Transportation Science*, Vol. 11, No. 1, 1977, pp. 60–72.
19. Treiber, M., and A. Kesting. *Traffic Flow Dynamics: Data, Models and Simulation*. Springer, Berlin, 2013.

TRAFFIC FLOW MODELS

Data Fusion Solutions to Compute Performance Measures for Urban Arterials

SERGE P. HOOGENDOORN

HANS VAN LINT

Delft University of Technology, the Netherlands

ROBERT L. BERTINI

California Polytechnic State University, San Luis Obispo

One of the key problems faced by traffic management operators of large urban traffic networks is the lack of sufficient data to compute performance indicators. These indicators, such as travel time, queue length, loss hours, and total time spent, are useful for offline evaluation purposes as well as online traffic control applications. In the latter case, such data are particularly of use in coordination algorithms that require information on the number of vehicles present or queuing in certain areas. This information in turn is used for example to assess the amount of buffer space available to temporarily store or reroute vehicles from more densely used parts of the network. Computing the amount of vehicles present or queuing in a certain area requires counting the number of vehicles that enter or exit that area. In this extended abstract the authors show how through fusing vehicle counts and travel times (measured by any means available), the well-known drift-error can be reduced to virtually zero. In the complete paper the authors show how this algorithm fits in a wider suite of data fusion tools to compute urban traffic performance indicators on the basis of multiple sources of data.

CUMULATIVE DRIFT

Denote $N(t)$ as the number of vehicles present on the road stretch at a certain time instant t . [Figure 1 \(top\)](#) illustrates that the number of vehicles on a simple hypothetical road stretch of 1 km can be derived directly using cumulative inflow and outflow curves $Q_1(t)$ and $Q_2(t)$. The horizontal distance between these curves depicts the travel time of the n th vehicle entering this road stretch, whereas the vertical distance depicts the number of vehicles $N(t) = Q_1(t) - Q_2(t)$. It is straightforward to compute $N(t)$ numerically (the authors consider discrete time steps k of size Δt):

$$N(k+1) = N(k) + \Delta t[q_1(k) - q_2(k)] \quad (1)$$

Unfortunately, this approach is highly sensitive to counting errors. Suppose both detectors make a 1% counting error that, for the sake of argument, is assumed normally distributed around the actual number of vehicles passing both locations:

$$q_i^{obs}(k) = q_i(k) + \eta(k) \quad (2)$$

with

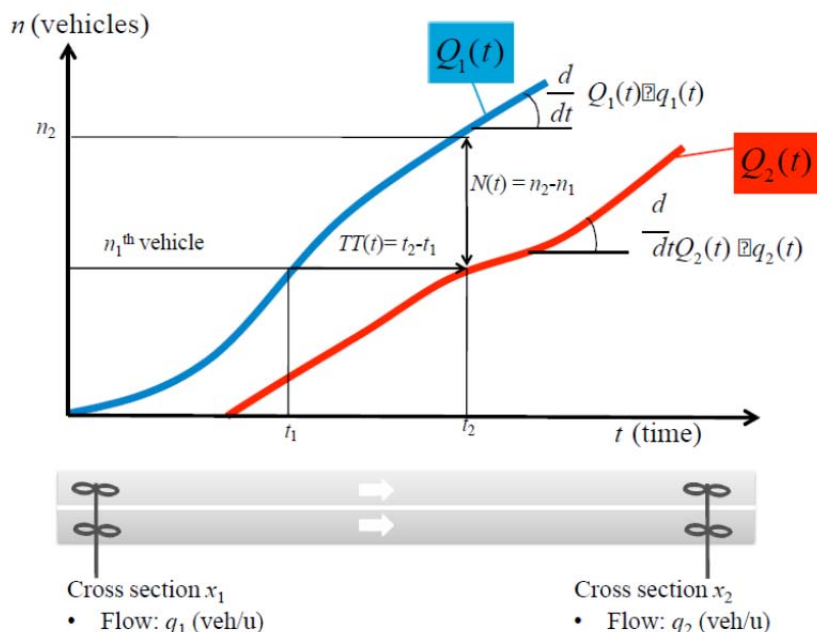


FIGURE 1 Cumulative inflow and outflow curves for a straight road stretch.

$$\eta_i(k) \sim N[0, 0.01 \times q_i(k)]$$

Substituting Equation 2 into Equation 1 results in Equation 3:

$$N(k+1) = N(k) + \Delta t[q_1(k) - q_2(k)] + \zeta(k) \quad (3)$$

with

$$\zeta(k) = \Delta t[\eta_1(k) - \eta_2(k)]$$

which describes a so-called random walk (the sum of two independent error terms) that may result in values anywhere between minus or plus infinity.

Figure 2a and b (top) show the measured in- and outflows $q_1(k)$ and $q_2(k)$ on the same hypothetical road stretch of 1 km, in which the errors of both 1% and 5% are hardly visible, whereas Figure 2a and b (bottom) show the resulting error in the estimated number of vehicles $N(k)$ using Equation 1. Even in case of a small error of 1% the error is sizeable—eight vehicles per km may make the difference between free-flowing traffic or oversaturated conditions.

FUSING FLOW OBSERVATIONS AND MEASURED TRAVEL TIMES

One approach to correct this so-called “drift” error is to use measured (realized) travel times $TT_{\text{obs}}(t)$ along the same road stretch. The underlying assumption is that travel time measurements are more accurate than loop counts (see Endnote 1). Recall that the horizontal distance between the cumulative curves allows researchers to estimate this travel time of the n th vehicle entering this road stretch.

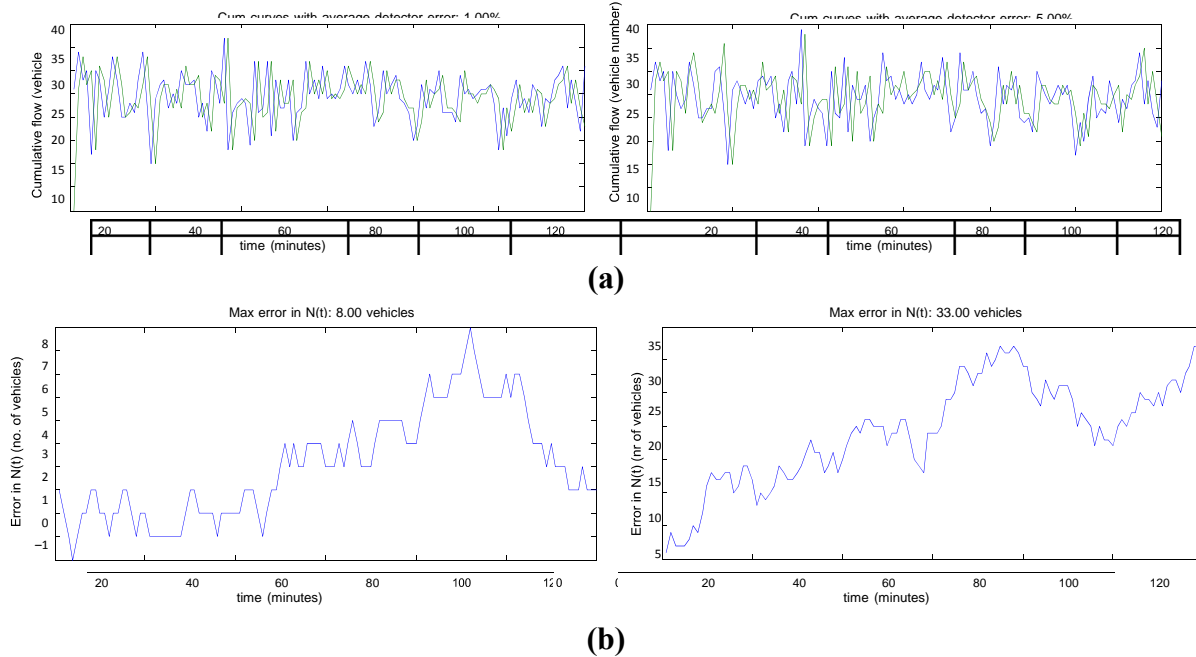


FIGURE 2 Cumulative drift: (a) measured in- and outflow flows $q_1(k)$ and $q_2(k)$ for detectors with a 1% and 5% error rate, respectively; (b) resulting errors in the estimated number of vehicles $N(k)$.

Figure 3 illustrates the basic idea. Suppose at some time instant t_2 researchers obtain a “travel time measurement” $TT_r^{\text{obs}}(t_2)$. At that same time instant an estimate for the realized travel time using the cumulative curves can also be computed:

$$TT_r^{\text{est}}(t_2) = t_2 - t_1$$

The error in this travel time estimate can now be used

$$\epsilon_{TT}(t_2) = TT_r^{\text{obs}}(t_2) - TT_r^{\text{est}}(t_2)$$

to correct both cumulative curves. Researchers define n^* as the factor with which to correct the curves. It turns out this correction factor is proportional to the travel time error, that is, $n^* \propto \epsilon_{TT}$.

This is true under the condition that outliers are properly removed before computing the average (or median). An example of such an outlier removal technique is the following:

Remove all travel times observations in a certain time period further away from the median travel time than $\alpha \times TT^{75} - TT^{25}$, where TT^{XX} represents the XX th travel time percentile. The outlier removal procedure should be applied strictly for this purpose.

$\epsilon_{TT}(t_2)$ —details will be provided in the full paper. The rationale of course makes sense. If the outflow (Q_2) is overestimated, then the rate at which vehicles are able to depart from this road stretch is overestimated, and the result is that the travel time these vehicles have incurred is underestimated. The opposite occurs when the inflow (Q_1) is overestimated; in that case the number of vehicles accumulating on the road stretch is exaggerated, and as a result travel time is overestimated. In Figure 3 the direction of this correction n^* is indicated in case the observed

travel time is larger than the estimated travel time [$\epsilon_{TT}(t_2) > 0$]. It turns out that this approach works remarkably well and is capable of eradicating the error in $N(t)$ virtually completely, not just in case of random errors but also in case of structural errors (bias).

Figure 4 gives an example of both 5% structural and random errors—again on the hypothetical 1 km road stretch introduced earlier. Figure 4a shows a detail of the corrected cumulative curves, the erroneous cumulative curves, and the ground-truth cumulative curves. In absolute sense, the algorithm does correct the errors, but relative to one another the algorithm does, which is demonstrated by time series of the erroneous versus the corrected number of vehicles estimated in Figure 4b—the latter lies very close to the ground truth.

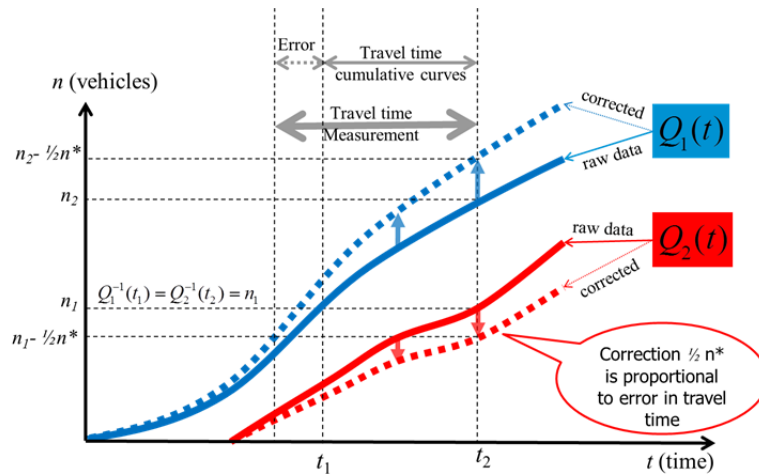


FIGURE 3 Basic idea of using measured travel times to correct for the cumulative errors that result from “drift.”

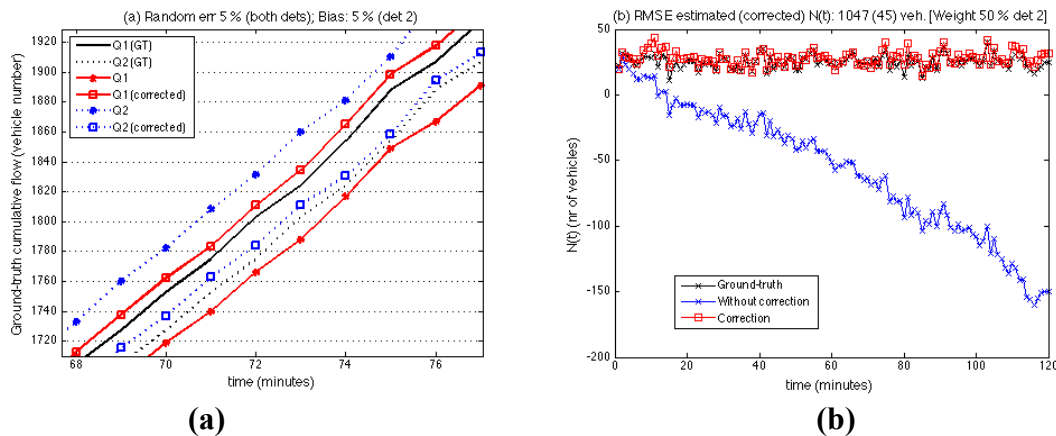


FIGURE 4 Example of error correction algorithm in case of random errors and structural errors: (a) a detail of the corrected cumulative curves, the erroneous cumulative curves, and the ground-truth cumulative curves; (b) time series of the erroneous versus the corrected number of vehicles estimated—the latter lies very close to the ground truth.

FULL PAPER OUTLOOK

This data fusion algorithm appears to fix cumulative drift errors by combining counts with travel times that may come from automatic vehicle identification systems, floating car data, Bluetooth sniffers, or any type of sensors. In the full paper the authors will discuss the algorithm in depth and demonstrate it on various real-life examples in the Netherlands and the United States.

ENDNOTE

1. This is true under the condition that outliers are properly removed before computing the average (or median). An example of such an outlier removal technique is the following: remove all travel times observations in a certain time period further away from the median travel time than $\alpha \times TT^{75} - TT^{25}$ where TT^{XX} represents the XX th travel time percentile. The outlier removal procedure should be applied strictly for this purpose.

TRAFFIC FLOW MODELS

Modeling Acceleration Behavior in a Connected Environment**ALIREZA TALEBPOUR****HANI S. MAHMASSANI***Northwestern University*

Conconnected vehicles technology will provide drivers with information on the presence and behavior of other drivers in their vicinity. This information is intended to help drivers make safe and reliable decisions. It will also affect drivers' strategic and operational decisions, with the most impact on the operational decisions, including acceleration choice. From the modeling standpoint, however, capturing the effects of this additional information on drivers' decisions is a challenging task and requires a more thorough understanding of humans' decision-making processes.

Acceleration behavior has been studied extensively in the literature, and several models with varying levels of complexity have been introduced to capture the underlying processes. Unfortunately, most of these models are designed to capture driving behavior in the absence of communications. Their modeling capabilities are even more limited in a mixed environment where only a portion of the vehicles are equipped with the essential communication tools. This additional information motivates different behaviors in this mixed environment. The addition of autonomous vehicles could further contribute to the complexity in this environment. This paper is intended to introduce an acceleration framework to capture the impacts of this additional information on driving behavior. Accordingly, different acceleration models with different assumptions are used for regular, connected, and autonomous vehicles.

ACCELERATION FRAMEWORK

This section provides an overview of the acceleration framework with a brief description of the acceleration models.

Modeling Vehicles with No Communication Capability

The drivers of these vehicles receive no information from other vehicles or from the traffic management center (TMC). They only get information from road signs (variable message signs and conventional signs). They also have a rough perception of other drivers' behavior in their vicinity. Moreover, their acceleration behavior has a probabilistic nature, and they are uncertain about other drivers' future behavior. This uncertainty may result in crash occurrence.

In general, drivers are seeking to travel at a desired speed while avoiding crashes. Avoiding crashes is an extremely important factor in drivers' decision making because of its severe consequences. Hamdar et al. (1) presented an acceleration model that avoids (most) crashes by specifying behavioral mechanism. An extension to this model was presented by Talebpour et al. (2), who recognized that drivers have different perceptions under congested versus uncongested regimes. Accordingly, they introduced two utility functions, one for

modeling driver behavior in congested regimes and one for modeling driver behavior in uncongested regimes. At each evaluation stage, based on drivers' perception of their surrounding traffic condition, drivers employ the corresponding value functions to evaluate the gains from the chosen acceleration. They introduced a binary probabilistic regime selection mechanism into the evaluation stage where drivers use the resulting utility to evaluate each acceleration value. Note that this study adopted Talebpour et al.'s acceleration framework to model car-following behavior in the absence of communication.

Modeling Communication-Ready Vehicles

These vehicles are expected to have the capability of sending and receiving information to and from other vehicles and infrastructure-based equipment. Assuming reliable connectivity in the vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications networks, each vehicle will receive information about other vehicles in this network. The driver also receives real-time updates about the TMC decisions (e.g., real-time changes in speed limit). However, this information may not be available at all times and locations, and drivers' behavior may change according to the amount of information they receive. Accordingly, four scenarios can be defined: active and inactive V2V communications and active and inactive V2I communications.

Active Vehicle-to-Vehicle Communications

Considering the flow of information in this V2V–V2I communications network, drivers are certain about other drivers' behaviors. Moreover, they are aware of driving environment, road condition, and weather condition downstream of their current location. Therefore, a deterministic acceleration modeling framework is suitable for modeling this environment. This paper uses the Intelligent Driver Model (IDM) (3) to model this connected environment. While capturing different congestion dynamics, this model provides greater realism than most of the deterministic acceleration modeling frameworks.

Inactive Vehicle-to-Vehicle Communications

In this driving environment, no active communication exists between vehicles. In case that V2I communications are unavailable, drivers' only sources of information are road signs and their perception of surrounding traffic condition. Drivers' behavior in this case can be modeled similarly to the case that vehicles have no communication capability. In the presence of V2I communications, drivers directly receive information about the TMC decisions. Drivers' behavior in this case can be modeled similarly to the case that vehicles have active V2I communications.

Active Vehicle-to-Infrastructure Communications

From the TMC point of view, active V2I communications will provide a basis to detect individual vehicle trajectories that can be used as high-precision input data to traffic control algorithms. From the driver's standpoint, V2I communications do not directly influence the drivers' acceleration choice. Therefore, the acceleration modeling approach under active V2I

communications depends on the availability of V2V communications. However, active V2I communications will provide real-time information about the TMC decisions (e.g., speed limit update in a speed harmonization system), which aim to improve safety and mobility.

Inactive Vehicle-to-Infrastructure Communications

In this driving environment, no direct communication exists between vehicles and the TMC. Without V2V communications, drivers' only sources of information are road signs and their perception of surrounding traffic condition. Drivers' behavior in this case can be modeled similarly to the case that vehicles have no communication capability. In the presence of V2V communications, drivers may receive information about the TMC decisions from other vehicles (if at least one vehicle in the V2V communications network receives information from the TMC). Drivers' behavior in this case can be modeled similarly to the case that vehicles have active V2I communications.

Modeling Autonomous Vehicles

Considering the ability of autonomous vehicles to constantly monitor other vehicles in their vicinity, an autonomous vehicle is certain about other drivers' behavior. Moreover, these vehicles can react instantaneously to any changes in the driving environment. Therefore, a deterministic acceleration modeling framework with reaction time set to zero is suitable for modeling this environment. This paper uses IDM (3) to model autonomous vehicles.

PRELIMINARY RESULTS

The preliminary simulations targeted a four-lane highway on the eastbound direction of I-290 near Chicago, Illinois (see [Figure 1](#)). This 3.5-mi-long segment has four on-ramps and three off-ramps, each with different characteristics and different merging length. An example of simulation results are presented in [Figure 2](#). This figure reveals the impact of market penetration rate on mobility. In this figure, it is assumed that reaction time of drivers will decrease by 50% in the presence of V2V communications. Note that V2I communications are inactive in this example. Based on this figure, higher penetration rate of connected vehicles results in higher breakdown flow (compare 2,000 veh/h at 0% penetration rate to 2,500 veh/h at 50% penetration rate) and density (compare 28 veh/km at 0% penetration rate to 38 veh/km at 50% penetration rate). Moreover, the scatter in the fundamental diagram decreases as the number of connected vehicles increases and at 100% penetration rate, the breakdown is completely eliminated. Therefore, it can be concluded that the efficiency of the highway increases as the number of connected vehicles increases.



FIGURE 1 Characterization of the selected segment in Chicago, Illinois.

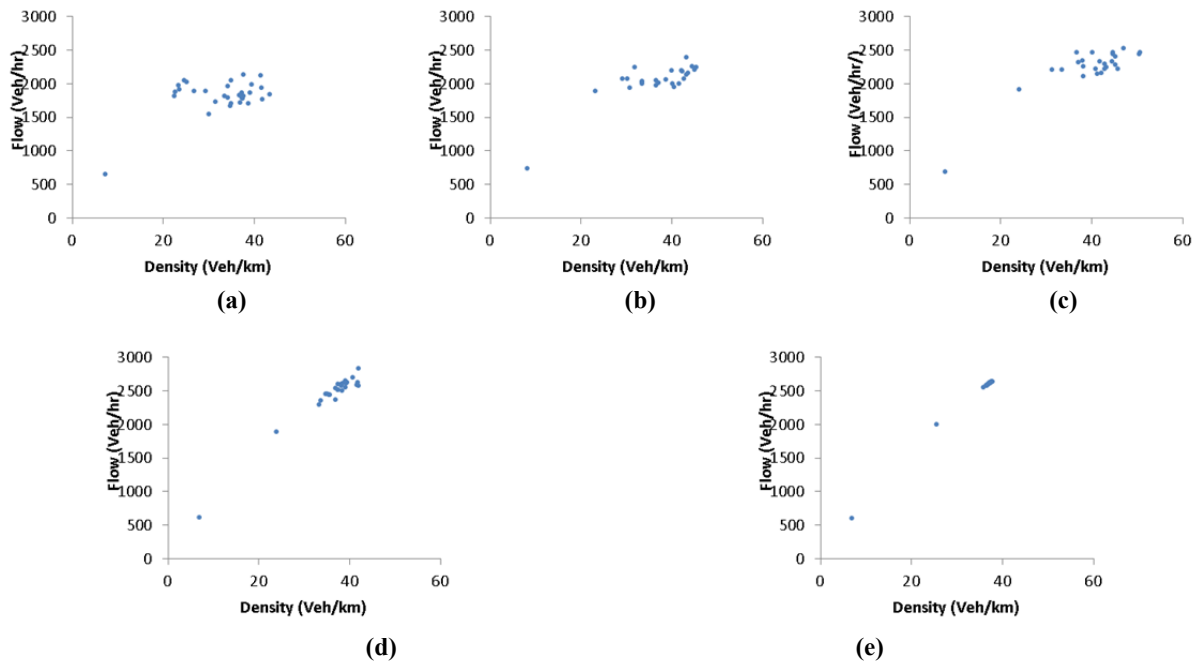


FIGURE 2 Flow–density graph for different market penetration rates: (a) 0%, (b) 25%, (c) 50%, (d) 75%, and (e) 100%.

REFERENCES

1. Hamdar, S. H., M. Treiber, H. S. Mahmassani, and A. Kesting. Modeling Driver Behavior as Sequential Risk-Taking Task. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2088, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 208–217.
2. Talebpour, A., H. S. Mahmassani, and S. H. Hamdar. Multiregime Sequential Risk-Taking Model of Car-Following Behavior. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2260, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 60–66.
3. Treiber, M., A. Hennecke, and D. Helbing. Congested Traffic States in Empirical Observations and Microscopic Simulations. *Physical Review E*, No. 62(2), 2000, pp. 1805–1824.

Traffic Control

TRAFFIC CONTROL

Traffic Flow Theory Milestones in Developing the TEXAS Model for Intersection Traffic in the Early 1970s

THOMAS W. (TOM) RIOUX

Rioux Engineering

The Traffic EXperimental and Analytical Simulation Model for Intersection Traffic (TEXAS Model) was developed by the Center for Transportation Research at the University of Texas at Austin beginning in the late 1960s under the leadership of Clyde E. Lee. Thomas W. Rioux was leader of the team of graduate students that developed the TEXAS Model and has been upgrading the TEXAS Model since its initial development. The TEXAS Model is being enhanced to include connected vehicle messages by Harmonia Holdings Group and Rioux to be a test bed for connected vehicle applications. The TEXAS Model source code is available for use by the public under the GNU General Public License as published by the Free Software Foundation. The TEXAS Model source code for the standard version may be downloaded from http://groups.yahoo.com/neo/groups/TEXAS_Model, while the version with connected vehicle applications may be downloaded from <http://www.etexascode.org>. This paper chronicles the evolution of the TEXAS Model from the early 1970s through 2008 and the early traffic flow theory concepts of triangular acceleration, triangular deceleration, equations of motion, car following, intersection conflict checking, intersection conflict avoidance, sight distance restriction checking, lane changing, and crashes.

INTRODUCTION

Microscopic traffic simulation involves defining the movement of individual driver–vehicle units through a roadway system in response to driver desires and control, other driver–vehicle units in the system, and the absence or presence of traffic control. A driver–vehicle unit is a vehicle with specified characteristics (such as type of vehicle, length, maximum acceleration, and maximum speed) controlled by a driver with specified characteristics (such as driver type, reaction time, and desired speed) that has an intersection origin leg and lane and a destination leg. Every driver–vehicle unit in the system is processed every small time-step increment (generally 1 s or less), wherein each individual driver makes many decisions (e.g., change lanes, slow down, speed up, stop, turn, avoid crash), vehicle detectors and signal controllers are simulated, and many measures of effectiveness (MOEs) are gathered and reported.

Clyde E. Lee was the faculty member who, in the late 1960s, conceived the idea of applying the University of Texas at Austin’s (UT) new Control Data Corporation (CDC) 1604 mainframe digital computer for simulating traffic flow through an intersection. He initiated the first development efforts and supervised several ensuing research projects that culminated in the TEXAS Model being released in 1977. Lee continued supervising research projects that enhanced or used the TEXAS Model until his retirement from UT in 1999. The TEXAS Model was developed by the Center for Highway Research and later the Center for Transportation Research (CTR) at UT using FORTRAN and mainframe computers. Initial funding for the

development efforts was provided by the Texas Department of Transportation (TxDOT) in cooperation with FHWA with later funding by FHWA and the UT College of Engineering.

The original TEXAS Model simulated a single intersection with no control, yield-sign control, less-than-all-way-stop-sign control, all-way-stop-sign control, pretimed-signal control, semi-actuated-signal control, or full-actuated-signal control using time-step increments between 0.5 and 1.5 s, inclusive, for a total of 4,500 s (1.25 h). The geometry included up to 6 legs with up to 6 inbound and 6 outbound lanes per leg; up to 1,000 ft straight lanes that could be blocked at the near end, at the far end, or in the middle; specification of movements allowed to be made from each inbound lane; specification of movements allowed to be accepted for each outbound lane; sight distance restrictions; detailed intersection path geometry using arcs of a circle and tangent sections; and the calculation of potential points of geometric conflicts between intersection paths. The traffic stream was stochastically generated using constant, Erlang, Gamma, lognormal, negative exponential, shifted negative exponential, and uniform distributions for headways with user-specified parameters; the normal distribution for desired speeds; and discrete percentages for turn movements, lane assignments, and other percentage-based parameters. For each inbound leg, the user specified the hourly volume, the headway distribution name and any parameters, the mean and 85th percentile speed, and the percentage of each vehicle class in the traffic stream. For each vehicle class (10 provided with a maximum of 15), the user specified the percentage of each driver class (three provided with a maximum of five).

The model included intersection conflict checking; sight distance restriction checking; cooperative lane changing using a cosine curve for the lateral position; car following using the Gazis-Herman-Rothery model ($I, 2$) with user-specified values for lambda (power for relative position), mu (power for speed), and alpha (constant); jerk-rate-driven equations of motion; triangular acceleration; triangular deceleration; and crashes with the driver-vehicle unit in front. MOEs included total delay (actual travel time minus the time to travel the same distance at the time-averaged desired speed), queue delay (time from initially joining the end of the queue of driver-vehicle units at the stop line until crossing the stop line), stopped delay (time stopped from initially joining the end of the queue of driver-vehicle units at the stop line until crossing the stop line), delay below a user-specified speed such as 10 mph, vehicle-miles of travel, travel time, volume, time and space mean speed, turn percentages, maximum and average queue length in 20-ft vehicles, and number of crashes. The MOEs could be printed per driver-vehicle unit and were summarized per lane or movement, per inbound leg, and for the entire intersection.

Initial model development effort began in 1968. Many students, faculty, and staff at UT have been involved in the development and use of the TEXAS Model:

- James W. Thomas, a graduate student in civil engineering at the time, began defining the concepts and techniques for modeling traffic flow through an intersection.
- Roger S. Walker, a graduate student in electrical engineering at the time, wrote some of the earliest CDC 1604 computer code for the TEXAS Model. His work included the development of the COordinated Logic Entity Attribute Simulation Environment program, which provided extremely efficient storage of model data and implemented an efficient means for processing logical binary networks. He was assisted by Dennis Banks.
- Thomas W. “Tom” Rioux, a graduate student in civil engineering at the time, started work on the project in 1971, followed up on Dr. Walker’s initial work, and became the leader of the team that developed the TEXAS Model into a viable tool for practical use in traffic

engineering and research using the CDC 6600 computer system until the TEXAS Model was released in 1977 (3–11). Rioux was the primary person who developed the field data analog-to-digital processing software that was used for model validation, DISFIT, GEOPRO, SIMPRO, the CDC 250 Display System version of DISPRO, SIMSTA, REMOVEC, REPLACEC, and gdvsim. He also participated in the development of DVPRO, the Intergraph UNIX X Windows version of DISPRO, the Java version of geoplot, and the Java version of dispro. In 1973, Rioux developed an animation on the CDC 250 Display System that was used during initial development efforts. Field measurements of queue delay using specifically designed recording devices were used to calibrate and validate the TEXAS Model at a 4-leg intersection with pretimed signal control in Austin, Texas.

- Charlie R. Copeland, Jr., an undergraduate and then a graduate student in civil engineering at the time, was part of the original development team and was the primary person who developed DVPRO and EMPRO. He also participated in the development of the field data analog-to-digital processing software, DISFIT, GDVDATA, GDVCONV, SIMDATA, SIMCONV, and SIMPRO.
- Robert F. “Bobby” Inman, an undergraduate student in mechanical engineering at the time, was part of the original development team and was the primary person who developed the field data collection hardware, GDVDATA, GDVCONV, SIMDATA, SIMCONV, DISPRE, and the DOS version of DISPRO. He also led the development effort of the Texas Diamond and National Electrical Manufacturers Association (NEMA) traffic signal controller simulators within SIMPRO. Harold Dalrymple assisted him in the development of the field data collection hardware.
- Ivar Fett, a graduate student in civil engineering at the time, was the person who collected and analyzed the field data and developed the original lane-changing geometry and decision models, developed the initial all-way-stop sign control logic, and developed the initial pretimed signal control logic for SIMPRO. He participated in the development of the car-following logic for SIMPRO.
- William P. Bulloch, a graduate student in civil engineering at the time, developed the initial acceleration, deceleration, and car-following models for SIMPRO.
- Elia King Jordan, a graduate student in civil engineering at the time, developed the initial version of DVPRO.
- Glenn E. Grayson, a graduate student in civil engineering at the time, assisted in the development of the actuated signal control logic for SIMPRO and supervised the field data collection and analysis, which was used to validate the TEXAS Model.
- Vivek S. Savur, a graduate student in civil engineering at the time, assisted in the field data collection and analysis and assisted in the development of GEOPRO.
- Scott Carter, a graduate student in civil engineering at the time, was the primary person who developed the Intergraph UNIX X Windows version of DISPRO.
- Moboluwaji “Bolu” Sanu, a graduate student in electrical and computing engineering at the time, was the primary person who developed the Java versions of geoplot and DISPRO. He later participated in the Small Business Innovative Research Projects performed by Rioux Engineering (12, 13).
- Zhonghui Ning participated in the development of gdvsim in the Small Business Innovative Research Projects performed by Rioux Engineering (12, 13).

Many research projects have used the TEXAS Model, and their results are documented

elsewhere. The original software programs proved to be a very robust and logically sound platform on which numerous evolutionary enhancements, revisions, and new features were subsequently added through additional projects at CTR and Rioux Engineering as the TEXAS Model migrated from batch mode on a mainframe computer to interactive mode on modern microcomputers:

- 1977/12/01 V1.00, initial release;
- 1983/08/01 V2.00, emissions processor added (14);
- 1985/11/01 V2.50, converted to run on the DOS operating system on a microcomputer using 16-Bit FORTRAN compilers, user-friendly interface added, and DOS animation added (15);
 - 1989/01/01 V3.00, diamond interchange geometry and TxDOT Figure 3, 4, 6, and 7 dual-ring actuated diamond signal controller added (16);
 - 1992/01/31 V3.10, replicate runs added, wide or narrow output selection added, left-turn pull-out option added, hesitation factor added, maximum number of loop detectors per lane increased from 3 to 6, blocked lane processing modified, intersection conflict avoidance added, and driver-vehicle unit delay for unsignalized lanes modified;
 - 1992/03/25 V3.11, intersection conflict avoidance error fixed, lane change errors fixed, and look ahead algorithms modified;
 - 1992/12/15 V3.12, converted to run on the Unix operating system on a workstation, headway distribution fitting processor added, geometry plotting processor added, simulation statistics processor added, UNIX X Window animation added, free U-turns at diamond interchange added, Dallas diamond signal controller phase numbering added, NEMA TS 1-1989 signal controller with volume-density operation added, replicate runs for specified number of runs added, replicate runs to specified statistical tolerance added, spreadsheet macros developed, car following modified, and many small enhancements to numerous algorithms (17);
 - 1993/11/23 V3.20, car following modified and NEMA controller errors fixed;
 - 1994/05/10 V3.21, lane change error fixed;
 - 1994/06/07 V3.22, NEMA and Texas diamond controller errors fixed;
 - 1996/02/28 V3.23, car-following logic modified;
 - 1998/09/21 V3.24, utility programs from80d.exe and to80d.exe added and Y2K-compliant modifications made;
 - 2000/08/03 V3.25, Java animation added;
 - 2003/08/29 V4.00, compiled using 32-bit FORTRAN compilers and initial vehicle messages added;
 - 2005/08/12 V5.00, Java user interface added; geometry plotting processor converted to Java; source code released under GNU general public license as published by the Free Software Foundation; increased number of driver types to nine; increased number of vehicle types to 99; classify detector added; modified logical binary networks to use type LOGICAL variables; added Vehicle Message System (VMS) messages for special driver-vehicle units: (1) forced go time and duration, (2) forced stop location and duration, (3) forced run red signal time and duration; changed minimum time-step increment to 0.01 s; converted all real variables to double precision; added VMS message types: (1) Driver DMS, (2) Driver IVDMS, and (3) Vehicle IVDMS; added VMS messages: (1) accelerate or decelerate to speed x using normal acceleration or deceleration, (2) accelerate or decelerate to speed x using maximum vehicle

acceleration or deceleration, (3) stop at the intersection stop line, (4) stop at location x , (5) stop immediately using maximum vehicle deceleration, (6) stop immediately using crash deceleration, (7) change lanes to the left, (8) change lanes to the right, (9) forced go, and (10) forced run the red signal; added VMS messages: (1) start time, (2) active time, (3) location (lane or intersection path and beginning and ending positions), (4) driver-vehicle unit number (0 = all), and (5) reaction time distributions and parameters; Surrogate Safety Assessment Methodology (SSAM) file support added; Linux version developed (12, 18);

- 2008/07/31 V6.00, all user interface software made Section 508 compliant; built-in help and tool tips added; displaying the sight distance restrictions added; displaying the detector geometry and activity added; Java application developed to automate the running of the TEXAS Model; total simulation time extended to 9999.99 s (2.777775 h); lane length extended to 4000 ft; Java application to display statistics from 1 run or replicate runs developed; stop on crash using crash deceleration and remain stopped option added; crashes between driver-vehicle units on different intersection paths added; automated the running of SSAM; attach and display orthorectified image file added; updated the NEMA traffic signal controller simulator to NEMA TS 2-2003; pedestrians added as they affect the operation and timing of the NEMA and hardware-in-the-loop traffic signal controllers; pedestrian signal operation added to animation; caused other driver-vehicle units to react to a crash; dilemma zone statistics added; time-varying traffic for two or more periods added; hardware-in-the-loop traffic signal controller added; additional vehicle attributes added to articulate vehicles; distracted driver VMS message added; an optional lane change before or after the intersection to move from behind a slower driver-vehicle unit added; and simulation of bicycles, emergency driver-vehicle units, and rail driver-vehicle units added (13);

- 2010, Small Business Innovative Research (SBIR) project topic 10.1-FH3, “Simulating Signal Phase and Timing with an Intersection Collision Avoidance Traffic Model,” adding Society of Automotive Engineers J2735 Basic Safety Message, Signal Phase and Timing Message (SPAT), and Map Data Message (MAP) awarded to Harmonia Holdings Group, LLC, Blacksburg, Virginia; Phase I completed; Phase II in progress; and

- 2011, SBIR project Topic 11.1-FH2, “Augmenting Inductive Loop Vehicle Sensor Data with SPAT and GrID (MAP) via Data Fusion,” adding National Transportation Communications for ITS Protocol 1202 vehicle detector, traffic signal controller parameter, and traffic signal display messages awarded to Harmonia Holdings Group, LLC, Blacksburg, Virginia; Phase I completed; Phase II in progress.

TEXAS MODEL TRAFFIC FLOW THEORY

The TEXAS Model defines the perception, identification, judgment, and reaction time (PIJR) as a user-specified parameter for each driver class in seconds. Typical values are 0.5 for aggressive drivers, 1.0 for average drivers, and 1.5 for slow drivers. Throughout the remainder of this document, several functions and constants are used as follows:

- ABS(A) = absolute value A,
- ACOS(A) = arccosine of A,
- COS(A) = cosine of A,
- DT = time step increment in seconds,

$\max(A,B)$ = maximum value of A and B, and
 PI = value for PI.

TRIANGULAR ACCELERATION

An investigation of existing acceleration models was undertaken in the early 1970s by Lee and Rioux, and it was found that the uniform acceleration model did not match observed behavior accurately when considered on a microscopic scale. Using a Chi-squared goodness-of-fit test, a best-fit uniform acceleration model was calculated and the results plotted (see Figure 1) along with observed data points (19). This figure illustrates that the uniform acceleration model computes velocities that are too low during initial acceleration and that result in the driver-vehicle unit's reaching desired velocity much sooner than it should. A linear acceleration model, which hypothesizes use of maximum acceleration when vehicular velocity is zero, zero acceleration at desired velocity, and a linear variation of acceleration over time, was investigated. Comparisons of this model with observed data (see Figure 1) indicate excellent agreement. This model also compared favorably with the non-uniform acceleration theory (20) used in describing the maximum available acceleration for the driver-vehicle unit.

This work lead to the development of the triangular acceleration model used in the TEXAS Model (see Figure 2). The author will use the term "jerk rate" to describe the rate of change of acceleration or deceleration over time and is usually in units of feet per second per second per second. Starting from a stopped condition, a driver-vehicle unit will use a maximum positive jerk rate until it reaches the maximum acceleration, then the driver-vehicle unit will use a negative jerk rate until the acceleration is zero at the driver-vehicle unit's desired speed. The maximum acceleration is defined by the driver-vehicle unit's desired speed and the maximum acceleration for the driver-vehicle unit.

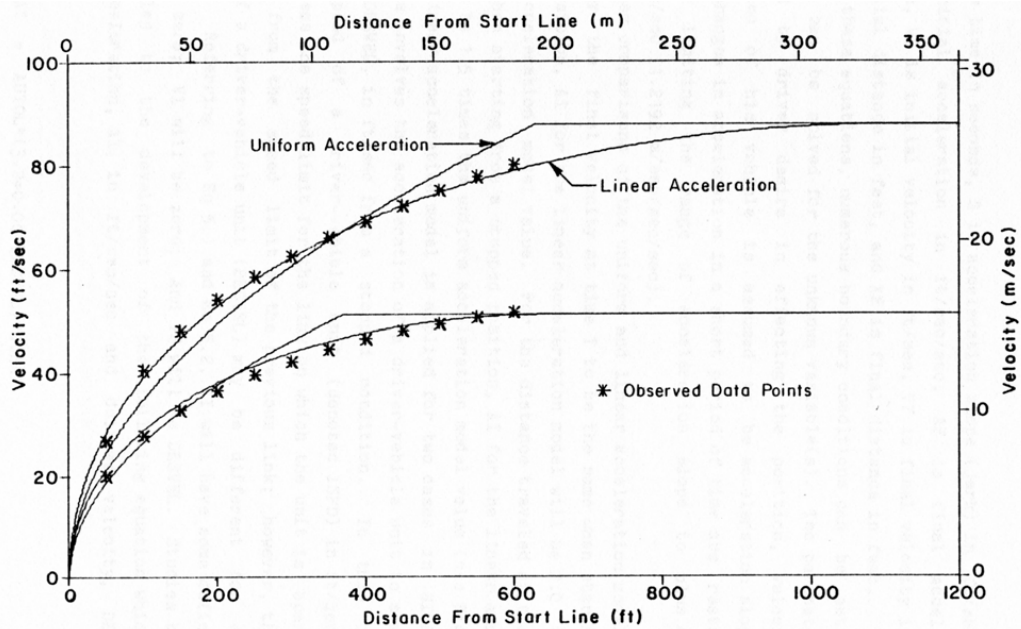


FIGURE 1 Uniform versus linear acceleration and observed data.

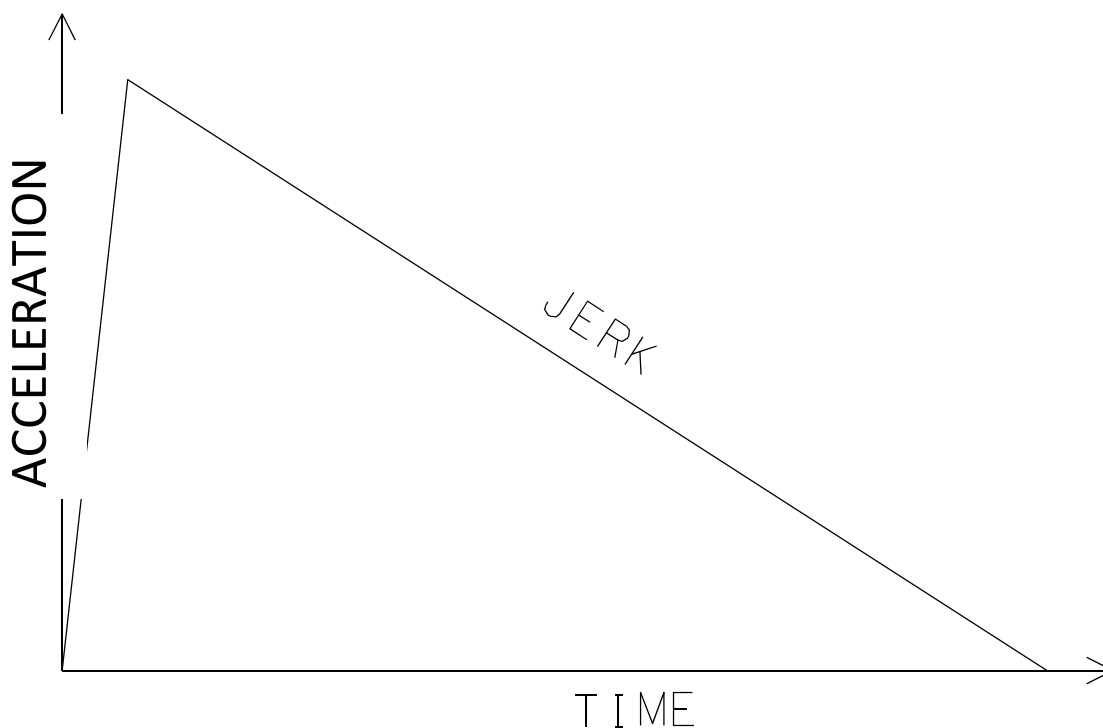


FIGURE 2 Triangular acceleration model.

TRIANGULAR DECELERATION

An investigation of existing deceleration models was also undertaken in the early 1970s by Lee and Rioux, and it was found that the uniform deceleration model did not match observed behavior accurately when considered on a microscopic scale. Using a Chi-squared goodness-of-fit test, a best-fit uniform deceleration model was calculated and the results plotted (see [Figure 3](#)) along with observed data points (19). This figure illustrates that the uniform deceleration model yields a higher velocity during the first part of the deceleration maneuver and, as the velocity approaches zero, produces values that are lower than observed values. A linear deceleration model, which hypothesizes use of a zero initial deceleration, maximum deceleration at the instant the driver-vehicle unit stops, and a linear variation of deceleration over time, was investigated. Comparisons of this model with observed data (see [Figure 3](#)) indicate excellent agreement.

This work led to the development of the triangular deceleration model used in the TEXAS Model (see [Figure 4](#)). Starting from a moving condition, a driver-vehicle unit will use a maximum negative jerk rate until it reaches the maximum deceleration when the driver-vehicle unit stops. The maximum deceleration is defined by the driver-vehicle unit's current speed and the maximum deceleration for the driver-vehicle unit. If a driver-vehicle unit is to decelerate to a stop, the time to stop and then the distance to stop is calculated each time step increment using current speed, current acceleration/deceleration, and current maximum deceleration. A deceleration to a stop is initiated when the driver-vehicle unit's distance to the location for a stop becomes less than or equal to the distance to stop.

EQUATIONS OF MOTION

With the development of the triangular acceleration and triangular deceleration models, it was clear that the equations of motion had to include jerk rate as follows:

$$AN = AO + J * DT$$

$$VN = VO + AO * DT + 1/2 * J * DT^2$$

$$PN = PO + VO * DT + 1/2 * AO * DT^2 + 1/6 * J * DT^3$$

where

AN = acceleration/deceleration new in feet per second per second,

AO = acceleration/deceleration old in feet per second per second,

J = jerk rate in feet per second per second per second,

PN = front bumper position new in feet,

PO = front bumper position old in feet,

VN = velocity new in feet per second, and

VO = velocity old in feet per second.

In the TEXAS Model, only the jerk rate is possibly changed each time step increment, and limits are placed on the maximum positive and negative values for jerk rate. Only in collisions are extremely large values of jerk rate used to stop a driver-vehicle unit in about 3-6 ft.

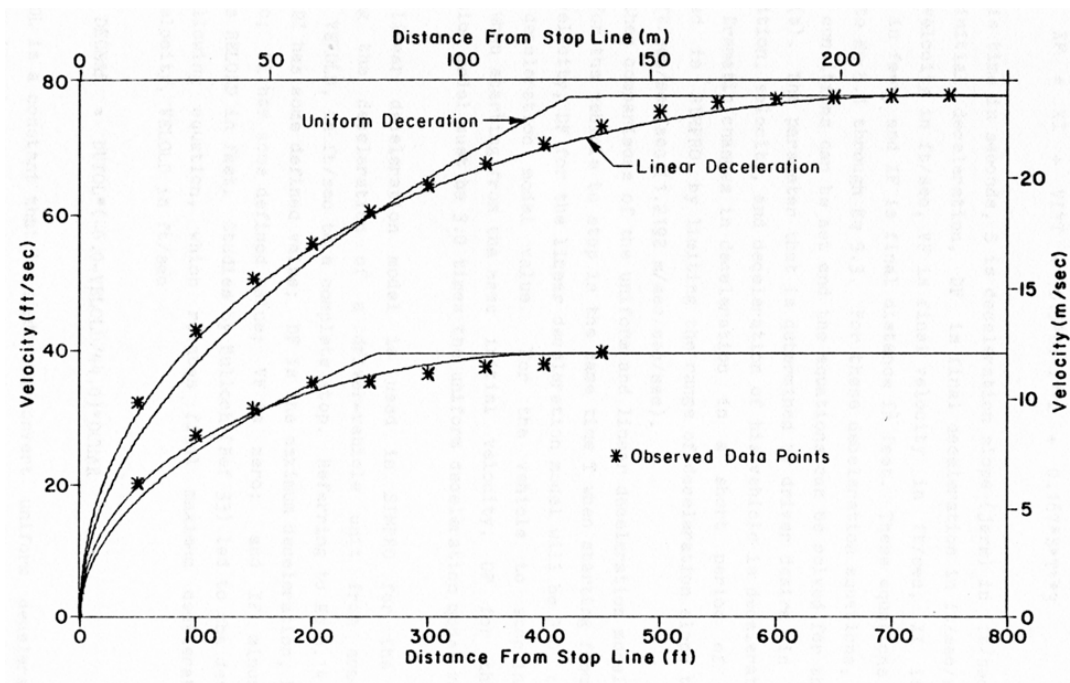


FIGURE 3 Uniform versus linear deceleration and observed data.

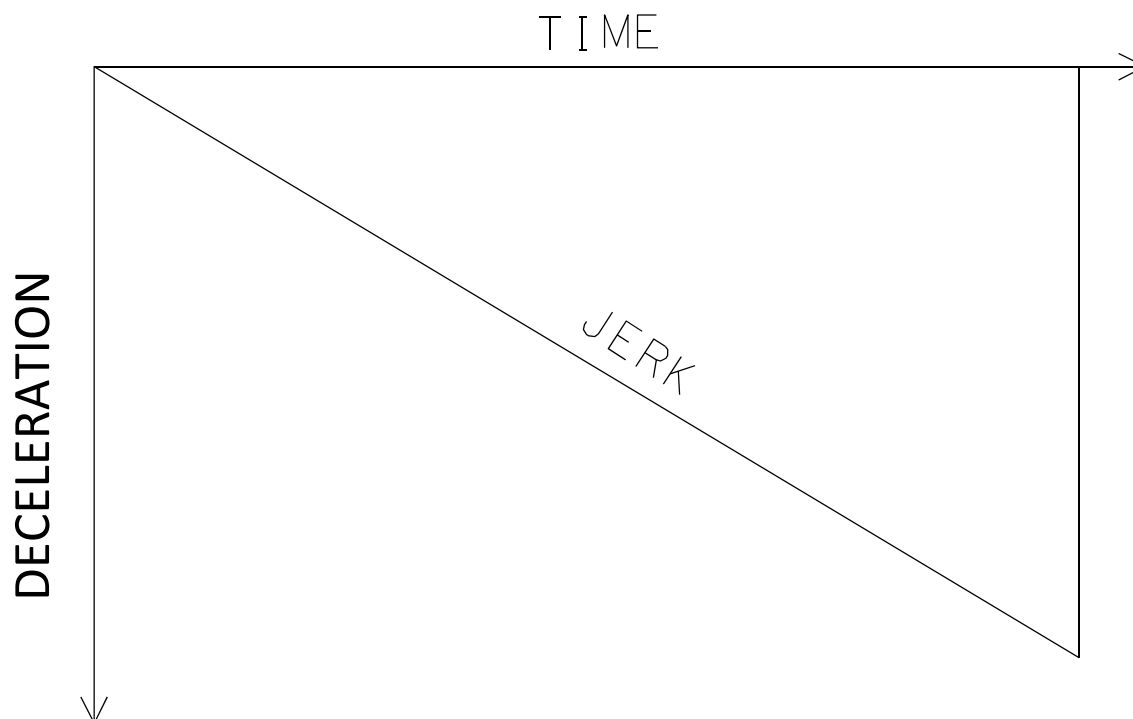


FIGURE 4 Triangular deceleration model.

CAR FOLLOWING

An investigation of existing car-following models was undertaken in the early 1970s by Lee and Rioux, and the non-integer, microscopic, generalized Gazis-Herman-Rothery (GHR) car-following model (1, 2) was selected because of its superiority and flexibility. If there is no previous driver-vehicle unit (no driver-vehicle unit ahead of the current driver-vehicle unit) then it cannot car follow, and thus other logic is used. If the previous driver-vehicle unit is stopped then it cannot car follow, and thus other logic is used. The GHR Model equation is as follows:

$$\begin{aligned} \text{RelPos} &= \text{PVPos} - \text{PO} \\ \text{RelVel} &= \text{PVVel} - \text{VO} \\ \text{AN} &= \text{CarEqA} * \text{VO}^{\text{CarEqM}} / \text{RelPos}^{\text{CarEqL}} * \text{RelVel} \end{aligned}$$

where

- AN = current driver-vehicle unit acceleration/deceleration new in feet per second per second;
- AO = acceleration/deceleration old in feet per second per second;
- CarEqA = user-specified GHR Model alpha parameter (min. = 1, def. = 4,000, max. = 10,000);
- CarEqL = user-specified GHR Model lambda parameter (min. = 2.3, def. = 2.8, max. = 4.0);
- CarEqM = user-specified GHR Model mu parameter (min. = 0.6, def. = 0.8, max. = 1.0);
- PO = current driver-vehicle unit front bumper current position old in feet;

PVPos = previous driver–vehicle unit rear bumper position in feet;
 PVVel = previous driver–vehicle unit velocity in feet per second;
 RelPos = relative position in feet;
 RelVel = relative velocity in feet per second; and
 VO = current driver–vehicle unit velocity old in feet per second.

The acceleration/deceleration new AN is not allowed to exceed the maximum acceleration or deceleration for the vehicle. The jerk rate to go from the current driver–vehicle unit acceleration/deceleration old AO to the current driver–vehicle unit acceleration/deceleration new AN is not allowed to exceed the maximum jerk rate. A conservative car-following distance is defined as follows:

$$\text{RelVel} = \text{PVVel} - \text{VO}$$

$$\text{CarDis} = (1.7 * \text{PVVel} + 4 * \text{RelVel}^2) / \text{DrivChar}$$

where

CarDis = car-following distance in feet;
 DrivChar = user-specified driver characteristic; (<1 = slow, 1 = average, >1 = aggressive, min. = 0.5, and max. = 1.5);
 PVVel = previous driver–vehicle unit velocity in feet per second;
 RelVel = relative velocity in feet per second; and
 VO = current driver–vehicle unit velocity old in feet per second.

If the relative velocity RelVel is greater than or equal to zero (the previous driver–vehicle unit is going faster than the current driver–vehicle unit) and the relative position RelPos is greater than some minimum value, then the driver–vehicle unit is allowed to accelerate to its desired speed.

If the relative position of the vehicle RelPos is less than or equal to zero, then emergency braking is applied. If the relative position of the vehicle RelPos is greater than the 1.2 times the car-following distance CarDis, then the driver–vehicle unit is allowed to accelerate to its desired speed.

If the previous driver–vehicle unit is decelerating, then calculate where it will stop and calculate the deceleration to stop behind the driver–vehicle unit ahead when it stops, and if this deceleration is less than the car following deceleration then use it.

If the traffic signal changed from green to yellow and the current driver–vehicle unit decides to stop on yellow, then calculate a deceleration to a stop at the stop line. If the traffic signal is yellow and the driver–vehicle unit previously decided to stop on yellow, then continue a deceleration to a stop at the stop line.

INTERSECTION CONFLICT CHECKING AND INTERSECTION CONFLICT AVOIDANCE

Intersection conflict checking (ICC) and intersection conflict avoidance (ICA) are essential algorithms for microscopic traffic simulation. ICC is the algorithm that determines whether a

driver-vehicle unit, seeking the right to enter the intersection, has a predicted time-space trajectory through the intersection that does not conflict with the predicted time-space trajectory through the intersection of all other driver-vehicle units that have the right to enter the intersection. ICA is the algorithm used to simulate the behavior of driver-vehicle units that have the right to enter the intersection and try to maintain a nonconflict time-space trajectory through the intersection with the predicted time-space trajectory through the intersection of other driver-vehicle units that have the right to enter the intersection.

Certain driver-vehicle units automatically gain the right to enter the intersection when there are no major collisions within the system: driver-vehicle units on an uncontrolled lane at a sign-controlled or signal-controlled intersection, driver-vehicle units going straight or right on intersection paths that do not change lanes within the intersection when the signal displays circular green, and all driver-vehicle units on signalized lanes when the signal displays protected green for their movement. Typical applications of ICC and ICA include a left-turning driver-vehicle unit crossing opposing leg straight through driver-vehicle units. The TEXAS Model included the ICC algorithm in Version 1.00 (released 12/01/1977), added the ICA algorithm in Version 3.10 (released 01/31/1992), and enhanced both algorithms in subsequent versions. The functionality and effectiveness of these algorithms have been verified extensively over the years by evaluation of the animation and analysis of the corresponding summary statistics from many varied simulations.

The TEXAS Model Geometry Processor (GEOPRO) calculates intersection paths starting at the coordinate for the middle of the stop line for an inbound lane, ending at the coordinate for the middle of the entry line for a diamond interchange internal inbound or outbound lane, tangent to the inbound lane, tangent to the outbound lane, and using the largest radius circular arc when needed. The user defines the turn movements that can be made from an inbound lane and the turn movements that can be accepted by an outbound lane. An intersection path consists of 4 segments in sequence. Each segment may or may not be used in the intersection path and is tangent at each end. The first segment is a tangent section, the second segment is an arc of a circle, the third segment is an arc of a circle, and the fourth segment is a tangent section. After calculating the geometry for all intersection paths, GEOPRO calculates the geometric conflicts between intersection paths including dual left turn side swipes (the intersection paths come within a user-specified distance but do not cross) and merges into the outbound lane. Finally, GEOPRO creates a list of geometric conflicts ordered by the distance from the beginning of the intersection path down the intersection path centerline to the point of geometric conflict. Data for each geometric conflict include the intersection path information and the conflict angle.

For each intersection path involved in a geometric conflict, the TEXAS Model Simulation Processor (SIMPRO) maintains a linked list of driver-vehicle units whose rear bumper plus a time safety zone has not crossed the point of geometric conflict. When a driver-vehicle unit gains the right to enter the intersection, SIMPRO adds the driver-vehicle unit to the end of the linked list for each geometric conflict for the driver-vehicle unit's intersection path. When a driver-vehicle unit is denied the right to enter the intersection, such as when a driver-vehicle unit decides to stop on a yellow signal indication, SIMPRO removes the driver-vehicle unit from the linked list for each geometric conflict for the driver-vehicle unit's intersection path. As the rear bumper plus a time safety zone crosses the point of geometric conflict, SIMPRO removes the driver-vehicle unit from the linked list for the geometric conflict for the driver-vehicle unit's intersection path.

To process the intersection conflicts for ICC for a driver-vehicle unit on an inbound lane

or diamond interchange internal inbound lane that has not gained the right to enter the intersection, SIMPRO first checks whether there are any geometric conflicts for the driver–vehicle unit’s intersection path, and if there are none, then intersection conflicts are clear. Next, SIMPRO processes each geometric conflict for the driver–vehicle unit’s intersection path in distance order. If a geometric conflict does not have a driver–vehicle unit whose rear bumper plus a time safety zone has not crossed the point of geometric conflict, then the geometric conflict is clear and the next geometric conflict is tested, else this geometric conflict is processed. In this discussion, “I,” “me,” or “my” refers to the driver–vehicle unit being processed, while “he,” “him,” or “his” refers to the next driver–vehicle unit whose rear bumper plus a time safety zone has not crossed the point of geometric conflict. The time for my front bumper to arrive at the geometric conflict (TCM), velocity at the geometric conflict for me (VCM), acceleration at the geometric conflict for me (ACM), and jerk rate at the geometric conflict for me (SCM) are predicted using my current distance to the geometric conflict, velocity, acceleration, jerk rate, driver characteristics, vehicle characteristics, speed limit for my intersection path, and information about any lead driver–vehicle unit that must be car-followed. The time for his front bumper to arrive at the geometric conflict (TCH), velocity at the geometric conflict for him (VCH), acceleration at the geometric conflict for him (ACH), and jerk rate at the geometric conflict for him (SCH) are predicted using his current distance to the geometric conflict, velocity, acceleration, jerk rate, driver characteristics, vehicle characteristics, speed limit for his intersection path, and information about any lead driver–vehicle unit that must be car-followed. A mini-simulation is used by SIMPRO to determine the time it takes the driver–vehicle unit to traverse the specified distance assuming that the driver–vehicle unit can accelerate to its desired speed or speed limit of its intersection path or car follow any lead driver–vehicle unit. The lead driver–vehicle unit, if any, is assumed to continue its current jerk rate. The velocity, acceleration, and jerk rate of the driver–vehicle unit when it has traversed the specified distance is also calculated. For ICC and ICA purposes, the lead gap is the space between my rear bumper and his front bumper when I go ahead of him through the geometric conflict, whereas the lag gap is the space between his rear bumper and my front bumper when I go behind him through the geometric conflict.

SIMPRO then calculates the time for the front safety zone for him (TFZ) and the time for the rear safety zone for him (TRZ) will arrive at the geometric conflict (see the top diagram in [Figure 5](#)) using the following equations:

$$\begin{aligned}
 \text{ERRJUD} &= \text{If } TCH > 5 \text{ then } \text{Max}(0.0, \text{PIJR} * (TCH - 5.0)/7.0) \text{ else } 0 \\
 \text{TPASSM} &= \text{LVAPM} / \text{VCM} \\
 \text{TPASCM} &= \text{DISCLM} / \text{VCM} \\
 \text{TPASSH} &= \text{LVAPH} / \text{VCH} \\
 \text{TPASCH} &= \text{DISCLH} / \text{VCH} \\
 \text{TFZ} &= TCH - \text{TPASSM} - \text{TPASCM} - (\text{TLEAD} - \text{APIJR}) - \text{PIJR} - \text{ERRJUD}/2 \\
 \text{TRZ} &= TCH + \text{TPASSH} + \text{TPASCH} + (\text{TLAG} - \text{APIJR}) + \text{PIJR} + \text{ERRJUD}/2 + \text{TPASCM}
 \end{aligned}$$

where

APIJR = average PIJR time for all driver–vehicle units in the entire traffic stream in seconds [calculated by the TEXAS Model Driver–Vehicle Processor (DVPRO)];

DISCLH = safety distance for him for merge into the same outbound lane in feet;
 DISCLM = safety distance for me for merge into the same outbound lane in feet;
 ERRJUD = error in judgment in seconds for TCH values greater than 5;
 LVAPH = length of vehicle along the intersection path for him at his current position in feet;
 LVAPM = length of vehicle along the intersection path for me at my current position in feet;
 PIJR = Perception, identification, judgment, and reaction time for the current driver–vehicle unit in seconds;
 TCH = time for his front bumper to arrive at the geometric conflict in seconds;
 TFZ = time for the front safety zone for him in seconds;
 TLAG = user-defined lag time gap for ICC in seconds (min. = 0.5, def. = 0.8, and max. = 3.0);
 TLEAD = user-defined lead time gap for ICC in seconds (min. = 0.5, def. = 0.8, and max. = 3.0);
 TPASCH = time for his driver–vehicle unit to pass through the geometric conflict because of a merge into the same outbound lane in seconds (zero if no merge);
 TPASCM = time for my driver–vehicle unit to pass through the geometric conflict because of a merge into the same outbound lane in seconds (zero if no merge);
 TPASSH = time for his driver–vehicle unit to pass through the geometric conflict in seconds;
 TPASSM = time for my driver–vehicle unit to pass through the geometric conflict in seconds;
 TRZ = time for the rear safety zone for him in seconds;
 VCH = velocity at the geometric conflict for him in feet per second; and
 VCM = velocity at the geometric conflict for me in feet per second.

The time period from TFZ until TRZ is blocked for me by his driver–vehicle unit. See the bottom diagram in Figure 5 to look at the time sequences from a gap perspective. If I can go safely in front of him (TCM is less than TFZ), or I can go safely behind him (TCM is greater than TRZ), then there is no conflict with his driver–vehicle unit at this geometric conflict. If I am blocked by his driver–vehicle unit at this geometric conflict (TCM is greater than or equal to TFZ, and TCM is less than or equal to TRZ), then there is a conflict with his driver–vehicle unit at this geometric conflict. If there is a conflict, then the ICC process is completed with a conflict found. If there is no conflict, I go behind him (TCM is greater than TFZ), and there is another driver–vehicle unit whose rear bumper plus a time safety zone has not crossed the point of geometric conflict, then I check the next driver–vehicle unit whose rear bumper plus a time safety zone has not crossed the point of geometric conflict. If there is no conflict and I go before him (TCM is less than or equal to TFZ), then I check the next geometric conflict for his intersection path because if I can go before him, then I can go before all other driver–vehicle units behind him. If all geometric conflicts for his intersection path have been checked and there are no conflicts, then the ICC process is completed with no conflict found. There are many special cases accommodated within the actual code when the geometric conflict is a merge, when there is a major collision somewhere within the system, when the other driver–vehicle unit is stopped and blocked by a major collision, when there is an emergency driver–vehicle unit in the system, or when a driver–vehicle unit is currently processing a forced go or forced run the red signal VMS message.

ICA is the algorithm used to simulate the behavior of driver–vehicle units that have the right to enter the intersection and try to maintain a nonconflict time–space trajectory through the intersection with the predicted time–space trajectory through the intersection of other driver–

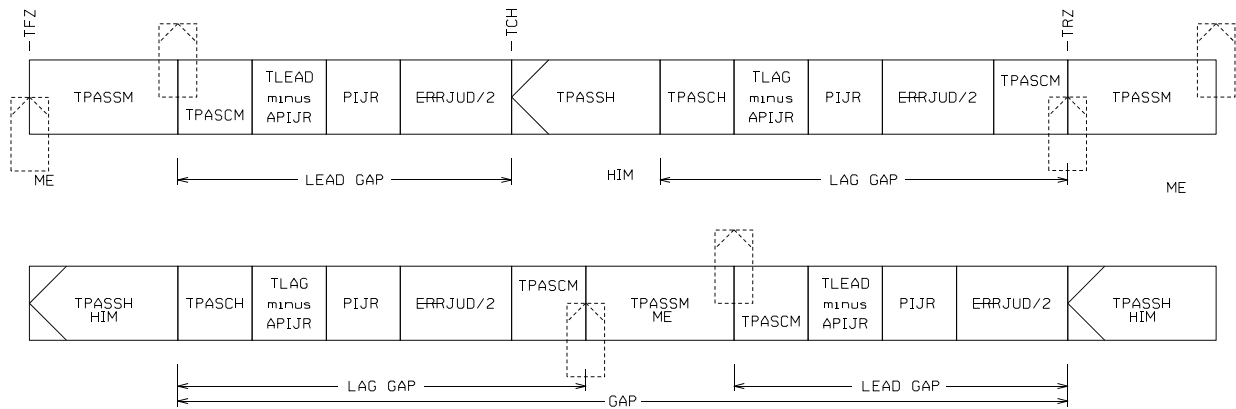


FIGURE 5 TEXAS Model ICC gap calculations.

vehicle units that have the right to enter the intersection. The linked list of driver-vehicle units whose rear bumper plus a time safety zone has not crossed the point of geometric conflict as described for ICC is also used for ICA. The jerk rate used for ICA (SLPCON) is initialized to 0.0.

To process the intersection conflicts for ICA for a driver-vehicle unit on an inbound lane or diamond interchange internal inbound lane that has gained the right to enter the intersection or a driver-vehicle unit that is within the intersection, SIMPRO uses a similar process as described for ICC. TCM, TCH, TFZ, TRZ, and the other variables are calculated in the same manner, and the same tests are performed to determine whether there is a conflict. The difference between the ICC and ICA processes is the action that is taken when a conflict is found. A variable TIM is calculated based on TCH, the turn movement for my intersection path, the turn movement for his intersection path, and whether there is a new green signal setting for me. TIM gives priority to a straight driver-vehicle unit over a turning driver-vehicle unit when they are both predicted to arrive at the geometric conflict at approximately the same time. If my turning movement is straight and his turning movement is straight, then TIM is set to TCH. If my turning movement is straight and his turning movement is left or right, then if I have a new green signal setting, then set TIM to TCH - 1.0, else set TIM to TCH + 1.5. If my turning movement is left or right and his turning movement is straight, then set TIM to TCH - 1.5. If my turning movement is left or right and his turning movement is left or right, then set TIM to TCH. Finally, if I am not an emergency driver-vehicle unit and he is an emergency driver-vehicle unit, then set TIM to TCH - 5.0. The jerk rate SLPTCM required for me to travel from my current position to the geometric conflict in time TCM starting with my current velocity and acceleration is calculated. This jerk rate represents the average value from the prediction process. If I have already passed the geometric conflict (TCM is less than or equal to 0.0), then nothing is done for this geometric conflict, and the next driver-vehicle unit or the next geometric conflict is processed.

The following logic is used when I am trying to go in front of him (TCM is less than or equal to TIM), therefore I try to accelerate to avoid the conflict. If the front safety zone for him has already arrived at the geometric conflict (TFZ is less than or equal to 0.0), then I should accelerate as fast as possible (set SLPTFZ to six times the critical jerk rate CRISLP). If the front safety zone for him has not already arrived at the geometric conflict (TFZ is greater than 0.0), then I should accelerate to go in front of him (set SLPTFZ to the jerk rate required for me to travel from my current position to the geometric conflict in time TFZ starting with my current velocity and acceleration). A temporary jerk rate SLPTMP is set to the maximum of (SLPTFZ -

SLPTCM) and 0.0. If I need to accelerate more than normal (SLPTMP is greater than 0.0), and there is no driver-vehicle unit ahead that I must car follow, and the temporary jerk rate is greater than the jerk rate used for ICA (SLPTMP is greater than SLPCON), then set SLPCON to SLPTMP. If I need to accelerate more than normal (SLPTMP is greater than 0.0), and there is a driver-vehicle unit ahead that I must car follow, and my speed is less than my desired speed, and the distance between me and the driver-vehicle unit ahead that I must car follow is greater than the car following distance, and the temporary jerk rate is greater than the jerk rate used for ICA (SLPTMP is greater than SLPCON), then set SLPCON to SLPTMP. The next driver-vehicle unit or the next geometric conflict is processed. This procedure will find the maximum positive jerk rate needed to accelerate to go in front of any driver-vehicle unit where a conflict has been found.

The following logic is used when I am trying to go behind him (TCM is greater than TIM), therefore I try to decelerate to avoid the conflict. If his rear safety zone has not reached the geometric conflict (TRZ is greater than 0.0), then I should decelerate to go behind him (set SLPTRZ to the jerk rate required for me to travel from my current position to the geometric conflict in time TRZ starting with my current velocity and acceleration). A temporary jerk rate SLPTMP is set to the minimum of $4.5 * (SLPTFZ - SLPTCM)$ and 0.0. If I need to decelerate more than normal (SLPTMP is less than 0.0), then set SLPCON to SLPTMP, and the ICA checking process is completed. This procedure will find the negative jerk rate needed to decelerate to go behind the first driver-vehicle unit where a conflict has been found. If SLPCON is not set to SLPTMP, then the next driver-vehicle unit or the next geometric conflict is processed.

If the jerk rate used for ICA has been set (SLPCON is not equal to 0.0), then SLPCON is added to the jerk rate calculated for this driver-vehicle unit (SLPNEW) if it is the critical value. There are many special cases accommodated within the actual code when the geometric conflict is a merge, when there is a major collision somewhere within the system, when the other driver-vehicle unit is stopped and blocked by a major collision, when there is an emergency driver-vehicle unit in the system, or when a driver-vehicle unit is currently processing a forced go or forced run the red signal VMS message.

SIGHT DISTANCE RESTRICTION CHECKING

The user defines the coordinates of all critical points needed to locate sight obstructions in the intersection area, and the TEXAS Model GEOPRO calculates the distance that is visible between pairs of inbound approaches for every 25-ft increment along each inbound approach. The TEXAS Model SIMPRO checks sight distance restrictions. Each driver-vehicle unit on an inbound approach assumes that it must stop at the stop line until it gains the right to enter the intersection. If the inbound lane is stop sign controlled or signal controlled, the assumption is made that sight distance restrictions are not critical and therefore do not need to be checked. If adequate sight distance is not available to a unit stopped at the stop line, this will not be detected in SIMPRO.

For driver-vehicle units on inbound lanes to an uncontrolled intersection, if there are units stopped at a stop line waiting to enter the intersection and the inbound driver-vehicle unit being examined is not stopped at the stop line, then the approaching driver-vehicle unit will continue to decelerate to a stop at the stop line without checking sight distance restrictions again

until it is stopped at the stop line or until there are no driver–vehicle units stopped at the stop line. This procedure eliminates unnecessary computations and gives the right-of-way to other driver–vehicle units already stopped at the stop line when the intersection is uncontrolled. If there are no sight distance restrictions for driver–vehicle units on an inbound approach then intersection conflicts are checked (see the ICC discussion above). If a driver–vehicle unit is on an uncontrolled lane approaching a yield-sign-controlled, if the driver–vehicle unit is stopped at the stop line, or if the intersection path of the driver–vehicle unit has no geometric intersection conflicts then it is assumed that there are no sight distance restrictions.

The maximum time from the end of the inbound lane that the driver–vehicle unit is permitted to begin checking sight distance restrictions, so that it may decide to proceed to ICC if sight distance restrictions are clear, is initially set to 3 s for all intersections. This prohibition prevents the driver–vehicle unit from gaining the right to enter the intersection when it is relatively far away from the intersection and thereby unnecessarily affecting the behavior of driver–vehicle units on other inbound approaches. If the inbound lane is an uncontrolled lane approaching a yield-sign-controlled intersection, then the time is increased by 2 s plus the time for the lead safety zone for ICC. This longer time allows driver–vehicle units on the uncontrolled lanes to gain the right to enter the intersection ahead of other driver–vehicle units on the yield-sign-controlled lanes. If the intersection is uncontrolled then the time is reduced to 2 s.

In SIMPRO, the time required for the driver–vehicle unit being checked to travel to the end of the lane is predicted. If this predicted time is greater than the maximum time from the end of the lane that the driver–vehicle unit may decide to proceed to ICC, then the driver–vehicle unit cannot clear its sight distance restrictions, and it must check again in the next time step increment.

The order in which sight distance restrictions are checked by SIMPRO is determined by the sequence in which intersection conflicts might occur. The sight distance restriction associated with the longest travel time to an intersection conflict is checked first, then other sight distance restrictions are checked in descending order of travel time to the intersection conflict. This order of checking facilitates early detection of an opportunity to pass in front of a driver–vehicle unit approaching on a sight-restricted lane. Checking continues until all inbound approaches that have possible sight distance restrictions with the subject inbound approach are cleared.

To check sight distance restrictions in SIMPRO, the time required for a fictitious driver–vehicle unit, traveling at the speed limit of the approach, to travel from a position that is just visible on the inbound approach to the point of intersection conflict is predicted. Next, the time required for the driver–vehicle unit being examined to travel to the point of intersection conflict is predicted. This prediction assumes that the driver–vehicle unit under examination has gained the right to enter the intersection and that it may accelerate to its desired speed. If the unit being checked may not safely pass through the point of intersection conflict ahead of the fictitious driver–vehicle unit, then it may not clear its sight distance restrictions, and it must check again in the next time step increment, otherwise, it clears the sight distance restriction and continues checking other sight distance restrictions.

This procedure ensures that a driver–vehicle unit may safely enter the intersection even if a driver–vehicle unit were to appear from behind the sight distance restriction just after the decision to enter the intersection was made.

LANE CHANGING

An investigation of lane changing models was undertaken in the early 1970s by Lee and Fett (5). Fett collected and analyzed the field data, developed the original lead and lag gap acceptance decision models, and used a cosine curve for the lateral position for a lane change.

Rioux developed the concept of distinguishing between two types of lane changes: (1) the forced lane change wherein the currently occupied lane does not provide an intersection path to the driver-vehicle unit's desired outbound approach, and (2) the optional lane change wherein less delay can be expected by changing to an adjacent lane that also connects to the driver-vehicle unit's desired outbound approach. Later, Rioux added cooperative lane changing and a lane change to get from behind a slower vehicle.

When a lane change is forced, a check is made to determine whether an alternate lane is geometrically available adjacent to the current position of the driver-vehicle unit being examined and is continuous to the intersection ahead. In the case of the alternate lane not being accessible from the current position, but available ahead, one of the two following conditions exists: (1) there is a lead driver-vehicle unit in the alternate lane ahead, in which case the driver-vehicle unit sets the lane change jerk rate to car follow the lead driver-vehicle unit in the alternate lane, or (2) there is not a lead driver-vehicle unit in the alternate lane ahead, in which case the lane change jerk rate is set to stop the driver-vehicle unit at the end of the alternate lane. If the end of the alternate lane has already been passed by the driver-vehicle unit when the check for an available alternate lane is made then the driver-vehicle unit is forced to choose one of the available intersection paths leading from the currently occupied lane and abandon the original destination. Otherwise, the driver-vehicle unit checks for an acceptable gap for lane changing.

When a lane change is optional, SIMPRO delays further lane-change checking until the driver-vehicle unit is dedicated to an intersection path. If there are no lane alternates adjacent to the current lane, then the lane change status flag is set to no longer consider a lane change. If the driver-vehicle unit is the first unit in the current lane and its intersection path does not change lanes within the intersection, then the lane change status flag is set to no longer consider a lane change. The expected delay is then computed for the driver-vehicle unit's current lane as well as for any alternate lanes. If less delay can be expected if the driver-vehicle unit changes into one of the alternate lanes, then that lane is checked for the presence of an acceptable lead gap and an acceptable lag gap, otherwise the process is repeated the next time step increment. If there is an acceptable lead gap and an acceptable lag gap, then the driver-vehicle unit is logged out of the current lane, logged into the new lane, and the lane change is initiated.

When the lead gap or the lag gap is not acceptable, the driver-vehicle unit tries to maneuver itself to make the gaps acceptable the next time step increment by accelerating, decelerating, or asking the lag driver-vehicle unit to car follow the current driver-vehicle unit to increase the lag gap (this is cooperative lane changing).

SIMPRO keeps track of the lateral position for the lane change old LatPosOld in feet which starts at the value for the total lateral distance for a lane change in feet TLDIST and decreases to zero when the lane change maneuver is completed. The lateral position of the lane change is computed using a cosine curve. Each time step increment, the current position on the cosine curve XOLD and the new position on the cosine curve XNEW are calculated as follows:

$$\begin{aligned} XTOT &= 3.5 * VO / (\text{DrivChar} * \text{VehChar}) \\ TLDIST &= 1/2 * \text{LanWidOrg} + 1/2 * \text{LanWidNew} \\ XOLD &= XTOT * \text{ACOS}(2 * \text{ABS}(\text{LatPosOld}) / TLDIST - 1) / \text{PI} \\ XNEW &= XOLD + VO * DT + 1/2 * AO * DT^2 + 1/6 * JN * DT^3 \end{aligned}$$

where

- AO = current driver-vehicle unit acceleration/deceleration old in feet per second per second;
- DrivChar = user-specified driver characteristic (<1 = slow, 1 = average, >1 = aggressive, min. = 0.5, and max. =1.5);
- JN = current driver-vehicle unit jerk rate new in feet per second per second per second;
- LanWidNew = new lane width in feet;
- LanWidOrg = original lane width in feet;
- LatPosOld = lateral position for the lane change old in feet;
- TLDIST = total lateral distance for a lane change in feet;
- VehChar = user-specified vehicle characteristic (<1.0 = sluggish, 1 = average, >1 = responsive, min. = 0.5, and max. =1.5);
- VO = current driver-vehicle unit velocity old in feet per second;
- XNEW = new position on the cosine curve in feet;
- XOLD = current position on the cosine curve in feet; and
- XTOT = total length of the lane change in feet.

If the new position on the cosine curve XNEW is greater than 95% of the total length of the lane change XTOT then the lane change is completed. The lateral position for the lane change new LatPosNew is calculated and stored as follows:

$$\text{LatPosNew} = 1/2 * TLDIST * [1 + \text{COS}(\text{PI} * XNEW / XTOT)]$$

where

- LatPosNew = lateral position for the lane change new in feet;
- TLDIST = total lateral distance for a lane change in feet calculated above;
- XNEW = new position on the cosine curve in feet calculated above; and
- XTOT = total length of the lane change in feet calculated above.

If lateral position for the lane change new LatPosNew is less than 0.3 ft, then the lane change is completed. Note that if the driver-vehicle unit speeds up, then the total length of the lane change XTOT increases, which causes the lane change to lengthen.

In 2008, Rioux extended the maximum lane length from 1,000 ft to 4,000 ft (13). This enhancement caused an additional optional lane change to be added before or after the intersection to move a driver-vehicle unit from behind a slower driver-vehicle unit. If the adjacent lane did not have an intersection path to the driver-vehicle unit's desired outbound approach, a lane change that would temporarily use the adjacent lane, pass the slower moving driver-vehicle unit, and lane change back into the original lane was performed if possible.

CRASHES

If the front bumper position of the driver–vehicle unit (lag driver–vehicle unit) is greater than the rear bumper position of the driver–vehicle unit ahead (lead driver–vehicle unit), then there is a crash. These were called “clear zone intrusions.” A message giving the details of the lead driver–vehicle unit and the lag driver–vehicle unit involved in the clear zone intrusion was output, and the clear zone intrusions were counted. The lag driver–vehicle unit defied physics by placing itself 3 feet behind the lead driver–vehicle unit traveling at the speed of the lead driver–vehicle unit and with zero acceleration/deceleration and jerk rate and the traffic simulation continued normally. Only crashes between a lead driver–vehicle unit and a lag driver–vehicle unit were detected.

In 2008, Rioux added the option to stop a driver–vehicle unit involved in a major crash using crash deceleration and remain stopped for the remainder of the simulation (13). This involved defining a major crash. Additionally, a crash between driver–vehicle units on different intersection paths was detected. Finally, code was added to cause other driver–vehicle units to react to driver–vehicle units involved in a major crash by slowing down as they passed near a crash if the driver–vehicle unit was not blocked by the major crash. After the driver–vehicle unit stopped because it was blocked by the major crash and a stochastically generated response time had elapsed, the driver–vehicle unit could possibly reverse a lane change maneuver if the driver–vehicle unit was still in the original lane or choosing a different intersection path to a possibly different desired outbound approach.

CONCLUSION

This paper chronicles the evolution of the TEXAS Model, which was developed by CTR at UT Austin beginning in the late 1960s. Topics include the early traffic flow theory concepts of triangular acceleration, triangular deceleration, equations of motion, car following, intersection conflict checking, intersection conflict avoidance, sight distance restriction checking, lane changing, and crashes. The TEXAS Model is being enhanced to include connected vehicle messages by Harmonia Holdings Group and for Rioux to be a test bed for connected vehicle applications.

The TEXAS Model source code is available for use by the public under the GNU General Public License as published by the Free Software Foundation. The source code for the TEXAS Model may be downloaded from

Standard version: http://groups.yahoo.com/neo/groups/TEXAS_Model
Version with messaging: <http://www.etexascode.org>

The TEXAS Model animations may be watched from YouTube (search YouTube for “TEXAS Model for Intersection Traffic Animation”):

1970s: <http://www.youtube.com/watch?v=1z4WleIOfbw>
1980s: <http://www.youtube.com/watch?v=S0utMJ9fZls>
1990s: <http://www.youtube.com/watch?v=PcU6WcaOAcE>
2000s: <http://www.youtube.com/watch?v=oah6nCGKwig>

Most of the references may be downloaded from files at http://groups.yahoo.com/neo/groups/TEXAS_Model_Documentation1

00000000_READ_ME.TXT
 00000001_TEXAS_Model_Development_History.txt
 19730126_TexITE.zip 19730500_Rioux_thesis.zip, z01, and z02
 19740500_Fett_thesis.zip 19770000_TRB_TRR_644.zip
 19771200_CTR_Research_Report_184-1.zip, z01, z02, z03, z04, and z05
 19771200_CTR_Research_Report_184-2.zip, z01, z02, z03, z04, z05, z06, and z07
 19770700_CTR_Research_Report_184-3.zip and z01
http://groups.yahoo.com/neo/groups/TEXAS_Model_Documentation2
 19771200_Rioux_dissertation.zip, z01, z02, z03, z04, z05, z06, z07, z08, z09, and z10
 19780700_CTR_Research_Report_184-4F.zip
 19801100_Torres_Evaluation_of_TEXAS_Model.zip
 19830800_CTR_Research_Report_250-1.zip, z01, z02, z03, z04, z05, z06, and z07
http://groups.yahoo.com/neo/groups/TEXAS_Model_Documentation3 19851100_CTR_Research_Report_361-1F.zip and z01 19890100_CTR_Research_Report_443-1F.zip, z01, z02, z03, and z04
 19910800_CTR_TEXAS_Model_Version_3_0_Documentation.zip, z01, z02, and z03
 19930100_CTR_Research_Report_1258-1F.pdf
 19931100_CTR_TEXAS_Model_Version_3_20_Documentation.zip, z01, and z02
 20040824_RiouxEngineering_DTRS57-04-C-10007_report.pdf
 20050800_CTR_DTFH61-03-C-00138.pdf
 20080731_RiouxEngineering_DTRT57-06-C-10016_report.pdf
 20100110 TRB Intersection Conflict Checking and Avoidance.pdf (not accepted)
 20120122 TRB Simulating Crashes and Creating SSAM Files.pdf
 20120122 TRB Simulating Crashes and Creating SSAM Files.ppt Evolution_of_Animation_of_the_TEXAS_Model.ppt TEXAS_Model_for_Intersection_Traffic.ppt TEXAS_Model_for_Intersection_Traffic_Section_508.ppt TEXAS_Model_Online_Documentation.htm

ACKNOWLEDGMENTS

The author thanks Clyde E. Lee, Professor Emeritus at the University of Texas at Austin, for his friendship, support, and guidance of the development of the TEXAS Model since 1971 and David Gibson, FHWA Turner–Fairbank Highway Research Center, for his friendship and support of the development of the TEXAS Model for many years.

The author also thanks the Texas Department of Transportation for its support of the TEXAS Model and the University of Texas at Austin Center for Transportation Research for allowing the source code for the TEXAS Model to be put into the public domain.

REFERENCES

1. Gazis, D. C., R. Herman, and R. W. Rothery. Nonlinear Follow-the-Leader Models of Traffic Flow. *Operations Research*, Vol. 9, No. 4, 1960, pp. 545–567.
2. May, A. D., Jr., and H. E. M. Keller. Non-Integer Car-Following Models. In *Highway Research*

- Record 199*, HRB, National Research Council, Washington, D.C., 1967, pp. 19–32.
3. Rioux, T. W. Step-Through Simulation Is Faster Than Driving. In *Compendium of the Annual Meeting of the Texas Section of the Institute of Traffic Engineers*, Bryan, Texas, January 26–27, 1973, pp. 54–73.
 4. Rioux, T. W. *Simulation of Traffic Movements in an Intersection*. MS thesis, University of Texas at Austin, 1973.
 5. Fett, I. *Simulation of Lane Change Maneuvers on Intersection Approaches*. MS thesis, University of Texas at Austin, 1974.
 6. Rioux, T. W. *The Development of the Texas Traffic and Intersection Simulation Package*. PhD dissertation, University of Texas at Austin, 1977.
 7. Rioux, T. W., and C. E. Lee. Microscopic Traffic Simulation Package for Isolated Intersections. In *Transportation Research Record 644*, TRB, National Research Council, Washington, D.C., 1977, pp. 45–51.
 8. Lee, C. E., T. W. Rioux, and C. R. Copeland. *The Texas Model for Intersection Traffic-Development*. Research Report No. 184-1, Project 3-18-72-184, Center for Highway Research, The University of Texas at Austin, 1977.
 9. Lee, C. E., T. W. Rioux, V. S. Savur, and C. R. Copeland. *The Texas Model for Intersection Traffic, Programmer's Guide*. Research Report No. 184-2, Project 3-18-72-184, Center for Highway Research, University of Texas at Austin, 1977.
 10. Lee, C. E., G. E. Grayson, C. R. Copeland, J. W. Miller, T. W. Rioux, and V. S. Savur. *The TEXAS Model for Intersection Traffic, User's Guide*. Research Report No. 184-3, Project 3-18-72-184, Center for Highway Research, University of Texas at Austin, 1977.
 11. Lee, C. E., V. S. Savur, and G. E. Grayson. *Application of the TEXAS Model for Analysis of Intersection Capacity and Evaluation of Traffic Control Warrants*. Research Report No. 184-4F, Project 3-18-72-184, Center for Highway Research, University of Texas at Austin, 1978.
 12. Rioux, T. W. *Enhancing the Usability of the TEXAS Model for Intersection Traffic Final Report*. Research Report Number SBIR DTRS57-04-C-10007-F, Federal Highway Administration Small Business Innovation Research Program Solicitation Number DTRS57-03-R-SBIR Contract Number DTRS57-04-C-10007, Rioux Engineering, Austin, Texas, 2004.
 13. Rioux, T. W., R. F. Inman, C. R. Copeland, Jr., M. Sanu, and Z. Ning. *Enhancing the Usability of the TEXAS Model for Intersection Traffic Final Report*. Research Report Number SBIR DTRT57-06-C-10016-F, Federal Highway Administration Small Business Innovation Research Program Solicitation Number DTRS57-03-R-SBIR Contract Number DTRT57-06-C-10016, Rioux Engineering, Austin, Texas, 2008.
 14. Lee, F.-P., C. E. Lee, R. B. Machemehl, and C. R. Copeland, Jr. *Simulation of Vehicle Emissions at Intersections*. Research Report No. 250-1, Project 2/3-8-79-250, Center for Transportation Research, University of Texas at Austin, 1983.
 15. Lee, C. E., R. F. Inman, and W. M. Sanders. *User-Friendly TEXAS Model, Guide to Data Entry*. Research Report No. 361-1F, Project 3-18-84-361, Center for Transportation Research, University of Texas at Austin, 1985.
 16. Lee, C. E., R. B. Machemehl, and W. M. Sanders. *TEXAS Model Version 3.0 (Diamond Interchanges)*. Research Report No. 443-1F, Project 3-18-84-443, Center for Transportation Research, University of Texas at Austin, 1989.
 17. Rioux, T., R. Inman, R. B. Machemehl, and C. E. Lee. *TEXAS Model for Intersection Traffic, Additional Features*. Research Report No. 1258-1F, Project 3-18-91/2-1258, Center for Transportation Research, University of Texas at Austin, 1993.
 18. Rioux, T. W. *Enhancement of the TEXAS Model for Simulating Intersection Collisions, Driver Interaction with Messaging, and ITS Sensors, Final Report*. Research Report Number DTFH61-03-C-00138, Center for Transportation Research, University of Texas at Austin, 2005.

19. Beakey, J. Acceleration and Deceleration Characteristics of Private Passenger Vehicles. In *Highway Research Board Proceedings*, Vol. 18, 1938, pp. 81–89.
20. Drew, R. R. *Traffic Flow Theory and Control*. McGraw-Hill, January 1968.

TRAFFIC CONTROL

Variable Speed Limit Control to Increase Discharge Rates of Freeway Incident Bottlenecks

DANJUE CHEN

SOYOUNG AHN

University of Wisconsin–Madison

Variable speed limit (VSL) schemes are developed based on the kinematic wave theory to increase discharge rates at freeway incident bottlenecks (BNs) while smoothing speed transition. The main control principle is to restrict upstream demand (in free flow) progressively to achieve three important objectives: 1. to provide gradual speed transition at the tail of an incident-induced queue, 2. to clear the queue around the BN, and 3. to discharge traffic at the stable maximum flow that can be sustained at the incident BN without breakdown. These control objectives are accomplished without imposing overly restrictive speed limits. This paper further provides remedies for the case of a reemerging queue at the BN as a result of an overestimated stable maximum flow. The results from a parameter analysis suggest that significant delay savings can be realized with the VSL control strategies.

INTRODUCTION

VSL control seeks to improve safety by smoothing out shockwaves at the tail end of a queue and freeway efficiency by deferring the onset of congestion or increasing the BN discharge rate. Earlier efforts to improve freeway efficiency using VSL focused on harmonizing the speed across vehicles in different lanes to create a more homogenous, stable one-pipe flow with few lane changes (LC) (1, 2). This can presumably lead to higher capacity and critical density, thereby deferring or preventing onset of congestion (3). Several studies have shown that VSL control indeed induces more balanced speed and utilization of lanes (1, 2, 4–6).

SPECIALIST (SPEEd ControllIng Algorithm using Shockwave Theory) seeks to proactively resolve a moving jam and maximize the discharge rate by limiting the speed and density of the inflow to the moving jam using VSL (7, 8). The algorithm was tested in the field with reasonable success. Another notable scheme is the mainstream traffic flow control (MTFC) developed in the framework of discrete-time optimal control (9). The main objective is to control the free-flow traffic upstream of a BN (before a queue arises) using VSL or ramp metering to prevent BN activation. The algorithm was tested on a Dutch network through simulation (10). Later, local feedback control was incorporated into MTFC to further improve field implementation (11).

Chen et al. (12) developed different VSL schemes based on the kinematic wave theory (13, 14) to increase freeway BN discharge rates and manage the queue upstream (for smoother speed transition) under two scenarios: steady queue and oscillatory queue that can inevitably arise at fixed BNs. The key principle is to impose VSL control some distance upstream of a BN to starve the inflow to the BN and dissipate the queue. Once the queue near the BN vanishes, another less restrictive VSL is imposed upstream to 1. resolve the heavy queue generated by the

first VSL and 2. regulate the inflow to sustain the stable maximum BN discharge rate and prevent BN re-activation.

The strategies cited above were designed primarily to address recurrent BNs or moving jams, in which a reduction in discharge rate typically ranges from 5% to 15% (15, 16). These strategies, however, are not suitable for nonrecurrent BNs such as incident BNs, in which discharge rates can reduce by more than 15%. Incident BNs are usually characterized by moderate to severe congestion caused by significant reductions in system throughput (in greater proportion than lanes blocked (17) and sharp transition upstream from free-flow traffic to the queue, which may cause secondary incidents. A significant reduction in system throughput is attributable to 1. a decrease in capacity because of road blockage, 2. rubbernecking around the incident, 3. change in driver characteristics (17), and 4. disruptive LCs away from incident location. With regard to 4, speed and flow may vary significantly among lanes with lower speed and flow closer to the incident location. LCs away from the incident location are likely to create voids in other lanes and reduce the discharge rate similar to the capacity drop phenomenon of recurrent BNs (18).

In this study, the authors develop VSL strategies to proactively improve the discharge rate of incident BNs and manage the upstream queue for smoother speed transition. An increase in discharge rate is achieved by clearing the queue around the incident and then maintaining a stable maximum flow with harmonized speed to minimize disruptive LCs. Note that other aforementioned factors for discharge rate reductions (1–3 in the previous paragraph) are not within the scope of this paper, because VSL control may be not the best option for these issues. The strategies developed in this paper are based on the kinematic wave theory and use the logic similar to Chen et al. (12). However, the new strategies address more effectively several critical issues for incident BNs: 1. A restrictive speed limit (lower than the speed in queue as in Chen et al. (12) should be avoided because incident-induced congestion is likely more severe; 2. the upstream queue management should be more elaborate to provide smoother queue transition; and 3. it may not be straightforward to precisely estimate the stable maximum discharge rate, which can lead to another queue formation at the incident BN.

This study develops a theoretical framework for VSL control to improve the performance of incident BNs. This is an important contribution given that existing strategies were primarily designed to mitigate recurrent BNs. The VSL control strategies developed in this study are relatively simple, yet capable of addressing two critical issues for incident BNs: 1. reducing incident-induced total delays significantly, and 2. providing smoother speed transition for better safety. The theoretical approach provides insights into traffic dynamics with VSL control and the impact of control parameters on the system performance (e.g., delay savings). Moreover, it provides a foundational framework to address more complex freeway networks and incorporate various implementation issues (e.g., detection and control technologies).

The remaining manuscript is organized as follows. Section 2 describes the basic VSL control strategy, including the analysis of parameters on the system performance and sensitivity. Two sequel VSL strategies are developed in Section 3 to remedy a reemerging queue at the incident BN caused by an overestimated stable maximum flow. Concluding remarks are provided in Section 4.

BASIC VSL CONTROL FOR INCIDENT BN

Baseline Case

This paper examines freeway BNs caused by incidents that may partially block the roadways and reduce the throughputs, as shown by Figure 1a. The authors assume that the traffic evolution can be well approximated by the kinematic wave model with triangular fundamental diagrams (FD). The upper FD in Figure 1b describes traffic states upstream of an incident with free-flow speed u , wave speed w , and jam density k_j ; and the lower FD describes traffic states at the incident location with lower free-flow speed u^{inc} and jam density k_j^{inc} . Note that it is assumed $u^{inc} < u$ because of rubbernecking and other effects induced by the incident.

The authors assume that traffic demand is constant in state A , and traffic breaks down to state H after the incident. After the incident is cleared, traffic recovers the full, normal capacity of q_{max} in state M . State e represents the stable maximum flow state that can be sustained at the incident BN without breakdown for an extended period (i.e., BN capacity, $q_{BN} = q_e$. The notation, q_{-} , represents flow in the traffic state denoted in subscript). Note that $q_H < q_{BN}$ because of LC disruptions and that $q_{BN} - q_H$ represents the potential gain of system throughput. States G and E correspond respectively to the free-flow and congested states with the same flow as q_{BN} (i.e., $q_{BN} = q_E = q_G = q_e$) upstream of the incident.

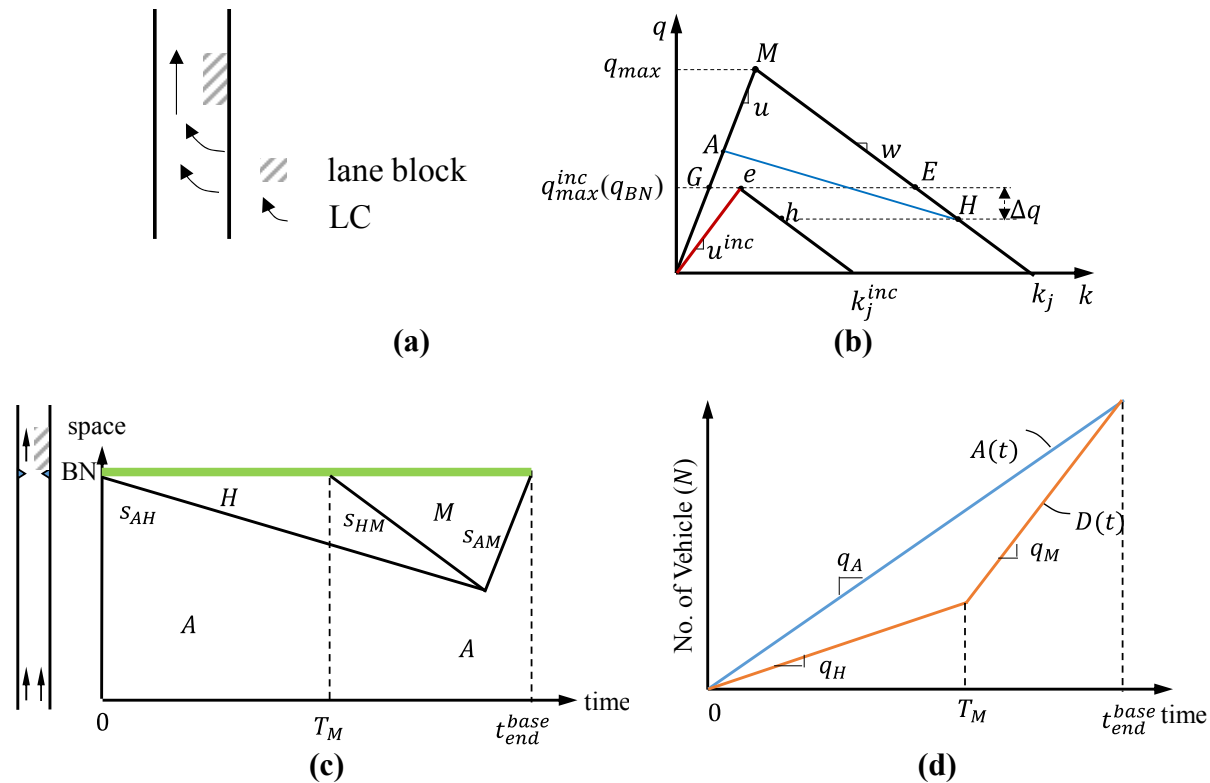


FIGURE 1 Traffic evolution at incident BN.

The spatio-temporal traffic evolution without any control is illustrated in Figure 1c. After the incident, a heavy queue (in state H) propagates upstream, forming a shockwave, s_{AH} . (Hereafter, s_{-} refers to a shockwave delineating two different traffic states denoted in subscript and represents shock wave speed when used in equations.) When the incident is cleared at time T_M , the normal capacity recovers and traffic evolves to state M , with the transition marked by s_{HM} . The queue ends when s_{AH} collides with s_{HM} , at which state A is resumed; this demand recovery is marked by s_{AM} . The queue ending time, t_{end}^{base} , and the total delay resulting from the incident, W^{base} , can be easily derived from the queuing diagram in Figure 1d, in which $A(t)$ and $D(t)$ denote the virtual arrival and departure curves at the BN, respectively:

$$t_{end}^{base} = \frac{(q_M - q_H)T_M}{q_M - q_A} \quad (1)$$

$$W^{base} = \int (A(t) - D(t))dt = \frac{1}{2}(q_A - q_H)T_M t_{end}^{base} = \frac{(q_A - q_H)(q_M - q_H)T_M^2}{2(q_M - q_A)} \quad (2)$$

Notably, if the speed in queued state H is low, which is quite likely with an incident, the speed drop along s_{AH} would be abrupt. In this case, it would be undesirable to impose a restrictive speed limit ($<v_H$) to clear the queue around the BN as prescribed by Chen et al. (12). Additionally, the upstream queue management strategies may not be sufficient to provide smooth enough transition at the queue's tail. A new VSL control strategy to address these problems is introduced below.

Basic VSL Control Strategy

The control principle is simple: restrict the upstream demand (in free flow) progressively. When done strategically, this will achieve three important objectives together: (1) to induce gradual transition at the queue's tail, (2) to clear the queue around the BN, and (3) to discharge traffic at stable maximum flow without breakdown. The control procedure consists of the following steps. (All traffic states resulting from the VSL control are shown on the FDs in Figure 2a, and the corresponding traffic evolution is shown in the time-space diagram in Figure 2b.)

Step 1

This main step is to control the upstream demand and clear the queue around the BN. Set several intermediate speed limit values between u and v_E with an even increment. (The notation, v_{-} , represents speed in the traffic state denoted in subscript.) The number of speed limit values will depend on the difference between u and v_E and the increment deemed acceptable to drivers. For an illustration purpose, the authors assume three intermediate values, denoted as V_1 , V_2 , and V_3 .

Step 1-1 At the start of control (t_0), impose V_1 simultaneously over an extended segment immediately upstream of the queue; see Figure 2b. The length of the segment will be discussed later. This results in a zone with state \tilde{a}_1 (Zone 1 in the figure), which has the same density as A , but the speed equals to V_1 ; see the FD in Figure 2a. Since the control is imposed simultaneously in space, the transition between A and \tilde{a}_1 forms a vertical shock, $s_{A\tilde{a}_1}$. At the downstream end of the control, state \tilde{a}_1 meets state H and forms $s_{\tilde{a}_1H}$. Notice that with this control, the queue propagates more slowly (i.e., $|s_{\tilde{a}_1H}| < |s_{AH}|$).

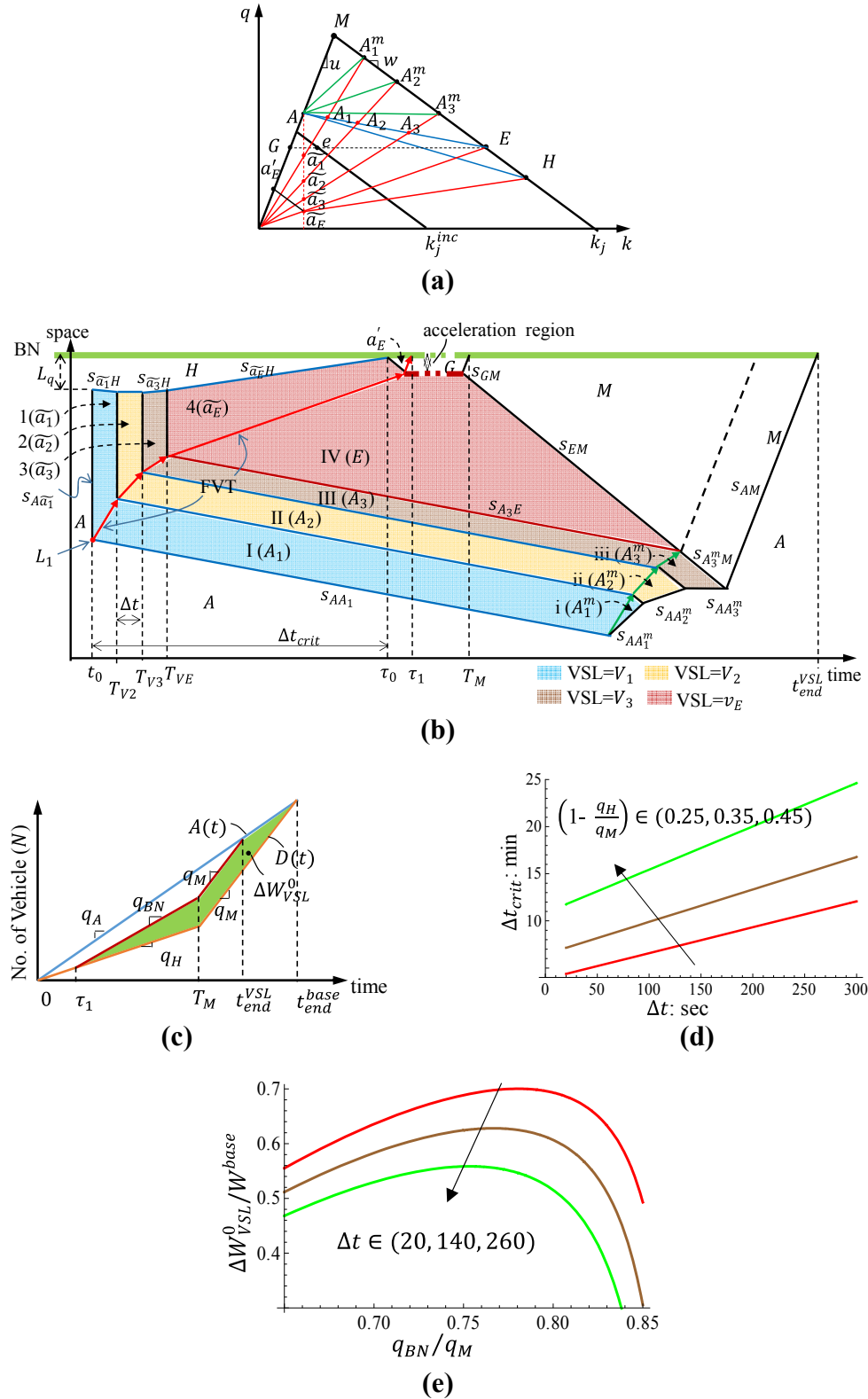


FIGURE 2 Basic VSL control strategy. (Note: On 2b the dotted line was used to indicate that the temporal duration of state G was shortened for presentation purposes to show the complete evolution. A similar way is used in Figure 3c and Figure 4a.)

Step 1-2 Switch the speed limit from V_1 to V_2 at time T_{V_2} to create Zone 2. Again, the new speed limit is actuated simultaneously in space immediately upstream of the queue. Similar to Zone 1, Zone 2 is in state \widetilde{a}_2 with the same density as A but with speed V_2 and forms $s_{\widetilde{a}_2H}$ where it meets state H . Notice that the queue propagates even more slowly (i.e., $|s_{\widetilde{a}_2H}| < |s_{\widetilde{a}_1H}| < |s_{AH}|$).

Step 1-3 Similar to Step 1-2, but change speed limit from V_2 to V_3 at time T_{V_3} to create Zone 3.

Step 1-4 Similar to Step 1-2, but change speed limit from V_3 to v_E at time T_{v_E} to create Zone 4 in state \widetilde{a}_E . Finally, the queue moves forward since $s_{\widetilde{a}_EH} > 0$ and is resolved $s_{\widetilde{a}_EH}$ arrives at the BN at τ_0 .

Step 2

This is to discharge traffic at the stable maximum flow without breakdown. The main idea is to have upstream traffic (in state A) gradually evolve to state E , and then let them fully accelerate to state G before passing the BN.

Create an acceleration zone immediately upstream of the BN so that traffic can accelerate to u . The length of this zone, L , is set to be 0.35-1 km as in Chen et al. (12).

Step 2-1 When H is resolved at τ_0 , deactivate v_E at the rate of the maximum backward moving wave speed, w , to guide the traffic in state \widetilde{a}_E to accelerate to state a'_E until the shock reaches the entrance of the acceleration zone; see Figure 2b. Note that $q_{a'_E}$ may be smaller than q_{BN} , resulting in underutilization of the BN capacity. Fortunately, this period is short ($= \left(\frac{L}{u} - \frac{L}{w}\right)$) and negligible compared to the incident duration. For example, it is about 4 min when $u = 100$ km/h, $w = -18$ km/h, and $L = 0.75$ km. Nevertheless, one possible remedy is to increase the deactivation rate to achieve a higher discharge rate than $q_{a'_E}$, which may be possible given the low density in \widetilde{a}_E (relative to the FD). For simplicity when calculating the delay saving, the authors assume $q_{a'_E} = q_H$, which corresponds to a deactivation rate, $w' \left(= \frac{q_H - (k_j + q_A/w)v_E}{q_H/u - (k_j + q_A/w)} \right)$.

The trajectory of the first vehicle that crosses the whole acceleration zone at u is denoted by the connected red arrows in Figure 2b, referred to as the first vehicle trajectory (FVT). The FVT serves as the boundary for the second set of VSL in the next step and is used to backcalculate the lengths of the vertical control at v_E , V_3 , V_2 , and V_1 , sequentially, as pictured in Figure 2b.

Step 2-2 Starting at the position of the FVT at t_0 , impose V_1 at the rate of s_{AA_1} ($s_{AA_1} = s_{AE}$). At the same rate, impose V_2 , V_3 , and v_E when the FVT intersects T_{V_2} , T_{V_3} , and T_{v_E} , respectively, such that all shocks are parallel. Speed limit v_E is extended to the entrance of the acceleration zone, at which drivers are informed to resume free-flow speed u (or v_e).

This step results in four zones: Zones I-IV, characterized by states $A_1 - A_3$ and E , respectively. Particularly, state A_1 in Zone I has the same speed as state \widetilde{a}_1 but a higher density; see Figure 2a. Notice that in Zones I through IV, vehicles gradually decrease their speed from u (free-flow speed) to v_E and then maintain v_E until reaching the entrance of acceleration zone.

Thereafter, they resume u and traffic evolves to state G . Notice that state G eventually evolves to state e at the BN without any flow change, as prescribed by the lower FD. (Recall that the lower free-flow speed at the BN is to capture rubbernecking and other effects of the incident.) In anticipation of this, traffic may alternatively be controlled to reach state e in the acceleration region. Regardless, the BN starts to discharge traffic at q_{BN} ($q_{BN} = q_E = q_G = q_e$) after the arrival of the FVT.

Step 2-3 When the incident is cleared at T_M , deactivate VSL to restore the full capacity of q_M . To restore q_M as soon as possible, deactivation at the entrance of the acceleration zone can be timed so that the forward moving shock, s_{GM} , will reach the BN at T_M . However, the gain is only about 1 min (the trip time of s_{GM}), and thus, one may safely start the deactivation at T_M . Concurrently, VSL is deactivated at the rate of w , forming s_{EM} , which eventually collides with s_{A_3E} and terminates Zone IV. Notably, the trajectory of the last vehicle that experiences state E is denoted by the three connected green arrows in Figure 2b, referred to as the last vehicle trajectory (LVT). As described later, the LVT will serve as the boundary for VSL deactivation.

Step 2-4 Deactivate V_1 , V_2 , and V_3 along the LVT to end the VSL control. As a result, three new zones, Zones i–iii, form naturally (i.e., no VSL control is needed) upstream of the LVT; see Figure 2b. States $A_1 - A_3$ evolve to congested states $A_1^m - A_3^m$, respectively, because of the bounded speeds, $V_1 - V_3$ downstream. When these states interact with the upstream demand, shocks $s_{AA_1^m} - s_{AA_3^m}$ form, which mark diminishing queue. As traffic emerges from state A_3^m , it evolves to the full capacity state, M , forming $s_{A_3^m M}$. When it collides with $s_{AA_3^m}$, the queue completely vanishes (Δt is in sec, $q_A = 0.8q_M$, $q_E = 0.75q_M$, $L_q = 2\text{km}$, $L = 0.75\text{km}$, $T_M = 2\text{h}$).

Parameter Analysis

The new queuing diagram with VSL control is shown in Figure 2c. Notice that the BN discharge rate increases to q_{BN} at τ_1 , resulting in quicker queue dissipation at t_{end}^{VSL} . τ_1 can be derived based on u and w' from the time–space diagram in Figure 2b, and t_{end}^{VSL} can be derived from the queuing diagram in Figure 2c.

$$\tau_1 = \tau_0 + L \left(\frac{1}{u} - \frac{1}{w'} \right) = -\frac{Lq}{s_{AH}} + \Delta t_{crit} + L \left(\frac{1}{u} - \frac{1}{w'} \right) \quad (3)$$

$$t_{end}^{VSL} = ((q_E - q_H)\tau_1 + (q_M - q_G)T_M) / (q_M - q_A) \quad (4)$$

where Δt_{crit} is the duration of queue clearance at the BN after VSL (total duration of Zones 1–4), as labeled in Figure 2b. Δt_{crit} is derived based on L_q , $s_{\bar{a}_1H}$, $s_{\bar{a}_2H}$, $s_{\bar{a}_3H}$, and $s_{\bar{a}_EH}$ and expressed as

$$\Delta t_{crit} = -\frac{3q_{AW}\Delta t(q_E(u-w)+k_jwu)+2L_q(q_E+k_jw)(q_Hu-q_{AW}+k_jwu)}{2w(-q_Eq_Hu+q_Aq_Ew-q_Hk_juw)} \quad (5)$$

There are several factors that affect Δt_{crit} and thus τ_1 , namely the durations of Zones 1–3, the number of intermediate VSL values, and the most restrictive VSL, v_E . For simplicity, it is assumed that the durations of Zones 1–3 all equal to Δt . Clearly, the sooner users start v_E (i.e., Δt is smaller), the sooner the queue at the incident BN can be cleared. Figure 2d illustrates the

impact, a linear positive relationship between Δt_{crit} and Δt , which increases faster when the congestion becomes more severe (i.e., q_H decreases as the capacity drop $(1 - \frac{q_H}{q_M})$ increases).

Note that Δt should be sufficiently long so that drivers can adapt and transition smoothly, which indicates the trade-off between the delay saving and speed transition. In the field implementation of Hegyi (19), such intermediate speed was set to last 20 s. Also notice that the number of intermediate VSL values between u and v_E can affect Δt_{crit} : given Δt , it will take longer to start v_E with more intermediate VSL values. This again highlights the trade-off between the delay saving and speed transition. Finally, the effect of v_E on Δt_{crit} is intuitive: increasing v_E increases Δt_{crit} and τ_1 since VSL becomes less restrictive. Conversely, increasing v_E can have a positive effect on $t_{\text{end}}^{\text{VSL}}$. As evident from Figure 2c, the BN would discharge traffic at a higher rate with higher v_E (recall that $q_E = q_{BN}$), increasing the rate of queue dissipation. This trade-off will be investigated in more detail shortly.

The delay saving as a result of VSL, ΔW_{VSL}^0 , equals to the area of the shaded region in Figure 2c.

$$\Delta W_{\text{VSL}}^0 = 0.5 * (q_{BN} - q_H)(T_M - \tau_1)((t_{\text{end}}^{\text{base}} - \tau_1) + t_{\text{end}}^{\text{VSL}} - T_M) \quad (6)$$

Figure 2e illustrates the impact of q_{BN} on ΔW_{VSL}^0 . Note that they are respectively expressed as the fractions relative to the road capacity and baseline total delay (i.e., q_{BN}/q_M versus $\Delta W_{\text{VSL}}^0/W^{\text{base}}$) to better examine the relative impact. It is interesting to note, however, that in severe congestion (e.g., $q_H < 0.5q_M$), the fractional delay saving increases with q_{BN}/q_M at decreasing rates up to a certain point and then decreases markedly. This trend is attributable to setting the most restrictive VSL at v_E that varies with q_{BN} , resulting in the trade-off between τ_1 and the rate of queue dissipation thereafter, as mentioned in the previous paragraph. More specifically, increasing v_E and q_E results in faster increases in τ_1 and thus decreases in the fractional delay saving. Also, notice that the impact of Δt is quite significant. For example, when Δt increases from 20 to 140 s, the delay saving can drop by more than 10%.

VSL CONTROL WITH REEMERGENCE OF QUEUE AT INCIDENT BN

Uncertainty in estimation of the stable maximum flow, q_{BN} , is a valid concern because of the nonrecurrent, wide-varying nature of incidents. Moreover, difficulty in predicting rubbernecking behavior adds to the challenge. In this section, the authors develop a VSL control strategy for the case of a reemerging queue at the incident BN, as a result of an overestimated q_{BN} .

It is assumed that the actual stable maximum flow, q_{BN}^* , is smaller than q_{BN} (and thus q_e , q_E and q_G); see Figure 3a. As a result, a queue forms again at the BN at τ_1 (when the BN is supposed to start discharging at q_{BN}), and traffic reverts to state H (see Endnote 1). This new queue propagates as s_{GH} , as depicted in Figure 3b. If no action is taken, the new queue would continue to travel upstream, albeit more slowly than the initial queue ($s_{GH} < s_{AH}$), and eventually terminate transition Zones I–IV. Moreover, the BN would discharge at q_H , meaning little to no savings in total delay.

To remedy this problem, new VSL control is imposed to clear the newly formed queue and then adjust the discharge flow. Two strategies, A and B, with different requirements of implementation, are proposed.

Strategy A

This is a simpler strategy with two steps, aiming at achieving q_{BN^*} as soon as possible. (All traffic states resulting from Strategy A are shown on the FDs in Figure 3a, and the corresponding traffic evolution is shown in the time–space diagram in Figure 3b.)

Step A-1

When a new queue is confirmed (at T_2), impose new VSL control with speed limit v_H simultaneously over a segment immediately upstream of the acceleration zone to clear the queue; see Figure 3b. By flow conservation, a new state, \tilde{E} , is created in Zone V with the same density as state E . The transition between E and \tilde{E} forms a vertical shock, $s_{E\tilde{E}}$. State \tilde{E} evolves to state I ($q_I = q_{\tilde{E}}$) in the acceleration zone and then resolves the queue, forming a forward moving shock, s_{IH} . The arrival of s_{IH} at the BN marks the clearance of the new queue, which indicates that q_{BN^*} at the BN can be resumed. The spatial extent of this VSL control at T_2 , denoted by $L_{\tilde{E}}$, should be designed so that state I does not persist at the BN (as pictured).

Upstream of Zone V, state E naturally evolves to state H (Zone VI) and forms s_{EH} . Note that state H in Zone VI is bounded by v_H because of state \tilde{E} downstream. Notably, s_{EH} is

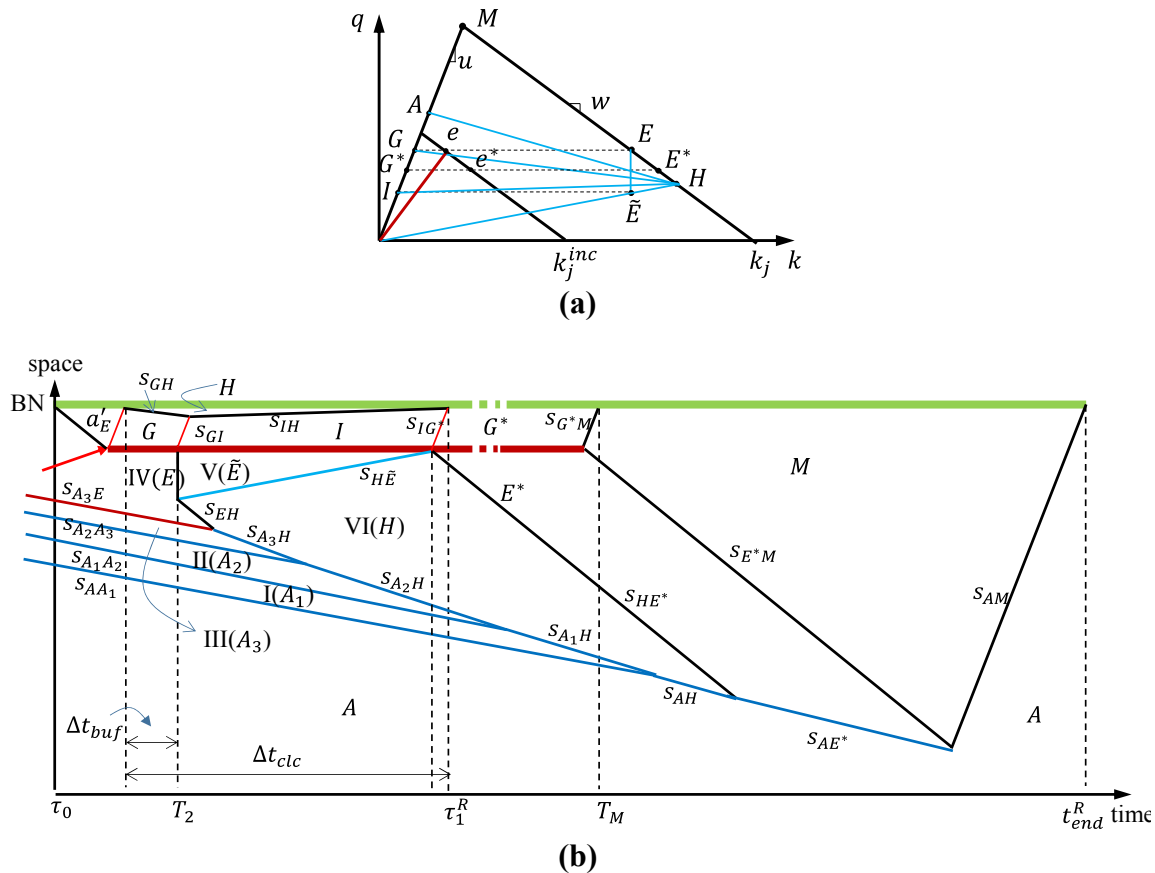


FIGURE 3 Strategy A to address queue reemergence.

terminated as it travels upstream and collides with s_{A_3E} . Thereafter, state A_3 (Zone III) interacts with H , forming s_{A_3H} . Similar transitions occur to A_2 (Zone II) and A_1 (Zone I) sequentially. In this process, the four transition zones (I–IV) are terminated by the queue in H gradually, and the speed transition between the intermediate states and H becomes gradually more abrupt. Eventually, state H interacts with state A directly.

Step A-2

When the queue at the BN is cleared, impose a less restrictive VSL, v_{E^*} , upstream of the acceleration zone to regulate the BN discharge rate at q_{E^*} ($= q_{BN^*}$). Upstream of the acceleration zone, traffic evolves from state H to E^* , forming s_{HE^*} , and then to state G^* in the acceleration zone. When the incident is cleared, this VSL control should be deactivated similar to Step 2-3.

State E^* propagates upstream until it is finally terminated by free-flow traffic A , marked by s_{AE^*} . After the VSL control is deactivated, traffic emerging from state E^* evolves to state M and resumes its full capacity.

Notably, this strategy is able to resolve the new queue and attain a higher BN discharge flow than q_H , yet at the expense of sharper speed transition upstream. Notably, transitions from A to E^* and from A to H are likely very abrupt in the absence of transition layers. Therefore, we introduce Strategy B to overcome this problem.

Strategy B

This strategy is built on Strategy A, but additional control upstream for smoother speed transition is added. Detailed steps follow. (All traffic states resulting from Strategy B are shown on the FDs in [Figure 4a](#), and the corresponding traffic evolution is shown in [Figure 4b](#).)

Step B-1

Same as Step A-1.

Step B-2

Same as Step A-2.

Step B-3

Notice in Strategy A ([Figure 3b](#)) that upstream speed transition is compromised because the heavy queue in state H (Zone VI) becomes widespread over time. In this step, the heavy queue will be contained to a shorter distance and resolved sooner by imposing VSL control on states A and $A_1 - A_3$ (transition zones I–III). Specifically, when the new queue is confirmed at T_2 , impose V_1 immediately upstream of state A_1 (Zone I) over the same spatial extent as A_1 (see Endnote 2). Following V_1 , impose V_2 , V_3 and v_{E^*} sequentially with time increment of Δt ; see [Figure 4](#), parts *a* and *b*. This creates low density and low flow states relative to the FD, $\tilde{a}_1 - \tilde{a}_3$ and \tilde{a}_{E^*} .

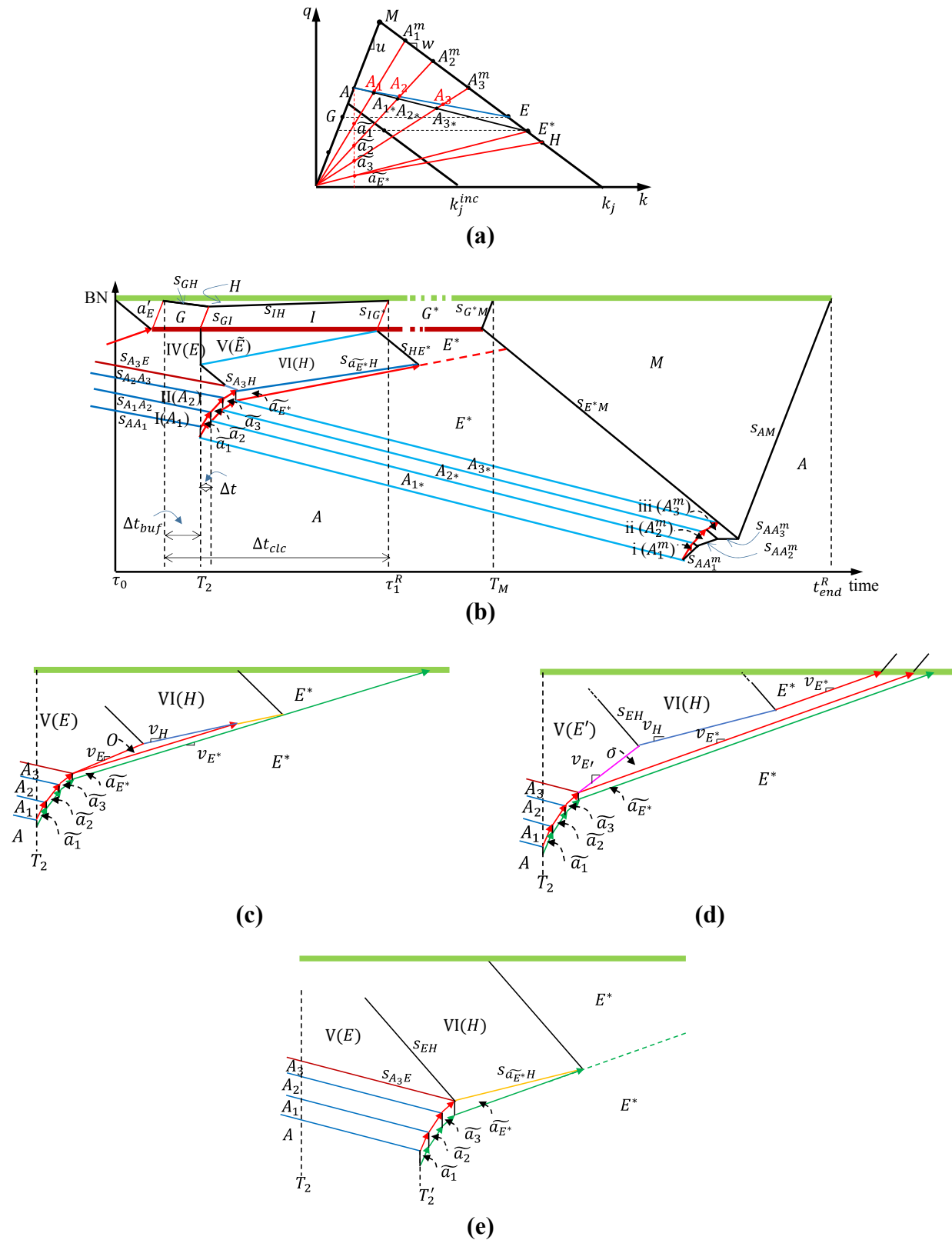


FIGURE 4 Strategy A to address queue reemergence.

The transition between \widetilde{a}_{E^*} and H forms a forward shock, $s_{\widetilde{a}_{E^*}H}$, which will contain and terminate state H when colliding with s_{HE^*} . The spatial extents of V_1 - V_3 and v_{E^*} at the start should be determined such that \widetilde{a}_{E^*} is terminated simultaneously with shock s_{HE^*} .

Note that state E may be terminated by state H early if s_{A_3E} and s_{EH} collide before v_{E^*} is supposed to go in effect, as depicted in Figure 4b. In this case, speed limit v_{E^*} is turned on sooner than Δt , more precisely, when $s_{A_3\widetilde{a}_3}$ collides with s_{A_3H} . By contrast, if state E still remains, speed limit v_{E^*} is imposed as planned (i.e., Δt later). This would create a void (state O) between \widetilde{a}_{E^*} and E (and later between \widetilde{a}_{E^*} and H) because traffic in E travels faster; see Figure 4c, which zooms in the region upstream of Zone VI in Figure 4b, to illustrate more detailed traffic evolution. If the speed difference between v_{E^*} and v_E is small, the void will eventually be resolved by \widetilde{a}_{E^*} . Thereafter, state \widetilde{a}_{E^*} proceeds to resolve Zone VI, marked by $s_{\widetilde{a}_{E^*}H}$. However, if the speed difference is significant, the void may pass through Zone VI and proceed to the BN, resulting in a discharge flow rate of 0, which is highly undesired; see Figure 4d. For this case, the upstream control (Step A-1) can be delayed until T'_2 , so that the three shocks, s_{EH} , s_{A_3E} , and $s_{A_3\widetilde{a}_3}$, converge; see Figure 4e. With this setting, researchers can actuate speed limit v_{E^*} following V_3 .

Step B-4

Upstream of \widetilde{a}_1 - \widetilde{a}_3 and \widetilde{a}_{E^*} , impose another set of V_1 - V_3 and v_{E^*} at the rate of s_{AE^*} to transition traffic gradually from A to E^* ($A \rightarrow A_{1^*} \rightarrow A_{2^*} \rightarrow A_{3^*} \rightarrow E^*$). Finally, deactivate the control when the incident is cleared, similar to Step 2.

Parameter Analysis

As expected, the delay saving would decrease if a queue reemerges at the incident BN; see Figure 5a for the queuing diagram. Note, however, that Strategies A and B have the same delay savings because the additional VSL control in Strategy B only manages the queue transition while leaving the BN discharge flow the same as in Strategy A. The delay saving (compared to the base case) is ΔW_{VSL}^R and can be derived based on the queuing diagram:

$$\Delta W_{VSL}^R = 0.5(q_{BN^*} - q_H)(T_M - \tau_1^R)((t_{end}^{base} - \tau_1^R) + t_{end}^R - T_M) \quad (8)$$

τ_1^R represents the time when q_{BN^*} is resumed and equals to $\tau_1 + \Delta t_{clc}$, where Δt_{clc} is the new queue clearance time; see the labels in Figure 3b.

Notice that Δt_{clc} depends on how responsive the system is. It is assumed that the new queue is detected after some buffer time, Δt_{buf} , or when the queue reaches the entrance of the acceleration zone, whichever happens first. Then, it is found that Δt_{clc} is given by

$$\Delta t_{clc} = \min \left(-\frac{k_j w(L + u \Delta t_{buf})}{q_H u}, -\frac{L(q_H(u-w) + u k_j w) k_j}{(q_E - q_H) q_H u} \right) \quad (9)$$

in which the former expression corresponds to early detection (i.e., after Δt_{buf}) and the latter for detection at the entrance. Notice that in the former case, a linear positive relationship exists between Δt_{clc} and Δt_{buf} . This is expected because the earlier the queue is detected, the sooner

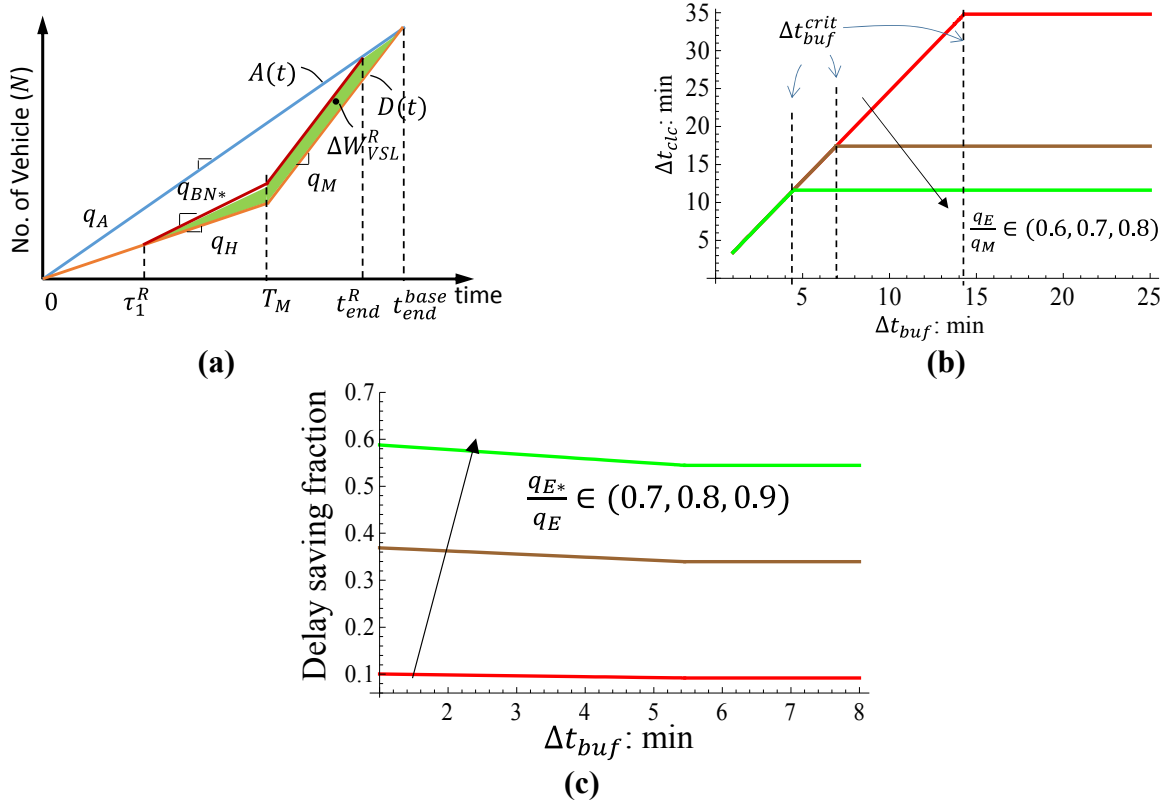


FIGURE 5 Parameter analysis of Strategies A and B
 [in (b – c), $q_A = 0.8 q_M$, $q_H = 0.5 q_M$, $L_q = 2$ km, $L = 0.75$ km, $T_M = 2$ h].

the queue clearance process can be actuated; see Figure 5b. In the latter case, Δt_{clc} depends on q_E , the initial estimation of the stable maximum flow. The higher the q_E is, the smaller the Δt_{clc} is. This is because a higher q_E (larger $|s_{GH}|$) indicates an earlier queue detection and faster queue dissipation (larger s_{IH}), both of which decrease Δt_{clc} . Notice that, the critical value of Δt_{buf} when the two cases are equal, denoted by Δt_{buf}^{crit} , decreases as q_E increases. Notably, this indicates that with a higher q_E , the queue (at the entrance) can be detected earlier and thus control actuated earlier. However, it also indicates shorter time available to set up the remedy action, which could pose a different challenge.

The relationship between the fractional delay saving ($\Delta W_{VSL}^R/W^{base}$) and Δt_{buf} is illustrated in Figure 5c. With early detection, it decreases as the buffer time increases, which is expected because Δt_{clc} is larger. Also notice that the delay saving depends on the new BN discharge rate, q_{BN^*} ($q_{BN^*} = q_{E^*}$): obviously, a larger q_{BN^*} leads to a higher delay saving.

Note that in Strategies A and B, speed limit v_H is used to clear the newly formed queue at the BN. This speed limit can be set to a lower value to clear the new queue faster. However, the minimum value, v_{min} , is bounded by the speed that yields $q_1 = q_H$:

$$v_{min} = \frac{q_H w}{q_E + k_j w} \quad (10)$$

A speed limit higher than v_{\min} would result in $s_{IH} < 0$, and it would not be possible to clear the newly formed queue.

Also note that in both strategies, the duration of Zone V (state \tilde{E}), $\Delta t_{\tilde{E}}$, is critical for the experiment set-up, which is given by

$$\Delta t_{\tilde{E}} = \Delta t_{clc} - \Delta t_{buf} - \frac{L}{u} = \min \left(-\frac{(L+u\Delta t_{buf})(q_H+k_jw)}{q_Hu}, \frac{L(q_H-k_jw)(q_H(u-w)+uk_jw)}{(q_E-q_H)q_Huw} \right) \quad (11-1)$$

Thereafter, the control distance at T_2 , $L_{\tilde{E}}$, can be derived accordingly:

$$L_{\tilde{E}} = \Delta t_{\tilde{E}}v_H = \min \left(-\frac{(L+u\Delta t_{buf})w}{u}, \frac{L(q_H-k_jw)(q_H(u-w)+uk_jw)}{(q_E-q_H)u(q_H+k_jw)} \right) \quad (11-2)$$

As revealed by Equation 11-1, in the case of early detection, $\Delta t_{\tilde{E}}$ (as well as $L_{\tilde{E}}$) increases with Δt_{buf} . This is because shock s_{IH} has to travel longer. For the late detection, both extensions increase as q_E decreases because the shock s_{IH} travels more slowly.

CONCLUSIONS AND DISCUSSIONS

In this paper, VSL strategies were developed based on the kinematic wave theory to increase discharge rates at freeway incident BNs and provide smoother speed transition upstream. The main logic is to impose VSL control gradually on upstream demand to dissipate the queue around the incident BN while inducing smoother speed transition at the queue's tail. After the queue clearance, VSL control continues to regulate inflow to the BN so that the BN can discharge traffic at the stable maximum rate in free-flow without breakdown. This is accomplished without imposing overly restrictive speed limits, an important feature considering incident situations. Findings from the parameter analysis suggest that significant delay savings can be realized with this strategy.

The authors further developed two sequel VSL strategies to remedy a reemerging queue at the incident BN because of an overestimated stable maximum flow. This is a likely scenario since incidents are nonrecurrent and vary widely in nature. The first, and the simplest, strategy is designed to clear the new queue and discharge traffic at the adjusted (lower) stable maximum rate, albeit with less desirable speed transition. The other two strategies were built on the first one to better manage the upstream queue for smoother transition. The two strategies differed by the magnitude of error in estimating the stable maximum flow. Not surprisingly, it was found that delay savings decreased when the stable maximum flow was overestimated; however, they were still substantial.

Building on the theoretical framework presented in this paper, ongoing research includes formulation of discrete schemes to 1. accommodate more complex scenarios, such as time varying demand and realistic freeway networks (with on and off-ramps), 2. shed light on spatiotemporal features of vehicle delay and speed variation, and 3. evaluate system robustness with respect to driver compliance and traffic detection/measurement errors, delays and resolution. Finally, the proposed VSL strategies should be tested in the field and further refined to address various practical issues, such as detection of traffic states and shockwaves.

ACKNOWLEDGMENTS

This research was sponsored by the National Science Foundation and the University of Wisconsin–Madison.

ENDNOTES

1. It is possible that a queue may form earlier if is sufficiently high. Note, however, that a VSL strategy to remedy this case would be similar.
2. This is because the vehicle trajectory passing these three zones is parallel to those passing Zones 1–3, such as the FVT.

REFERENCES

1. Smulders, S. Control of Freeway Traffic Flow by Variable Speed Signs. *Transportation Research Part B*, Vol. 24, No. 2, 1990, pp. 111–132.
2. Zackor, H. Speed Limitation on Freeways: Traffic-Responsive Strategies. In: Papageorgiou, M. (Ed.), *Concise Encyclopedia of Traffic and Transportation Systems*. Pergamon Press, 1991, pp. 507–511.
3. Papageorgiou, M., E. Kosmatopoulos, and I. Papamichail. Effects of Variable Speed Limits on Motorway Traffic Flow. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2047, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 37–48.
4. Duret, A., S. Ahn, and C. Buisson. Lane Flow Distribution on a Three-Lane Freeway: General Features and the Effects of Traffic Controls. *Transportation Research Part C*, Vol. 24, 2012, pp. 157–167.
5. Knoop, V. L., A. Duret, C. Buisson, and B. van Arem. Lane Distribution of Traffic near Merging Zones Influence of Variable Speed Limits. *Proc., 13th International IEEE Conference on Intelligent Transportation Systems*, Madeira Island, Portugal, 2010, pp. 485–490.
6. Weikl, S., K. Bogenberger, and R. L. Bertini. Traffic Management Effects of Variable Speed Limit System on a German Autobahn: Empirical Assessment Before and After System Implementation. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2380, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 48–60.
7. Hegyi, A., S. P. Hoogendoorn, M. Schreuder, and H. Stoelhorst. The Expected Effectivity of the Dynamic Speed Limit Algorithm SPECIALIST: A Field Data Evaluation Method. *Proc., European Control Conference*. Budapest, Hungary, 2009, pp. 1770–1775.
8. Hegyi, A., S. P. Hoogendoorn, M. Schreuder, H. Stoelhorst, and F. Viti. SPECIALIST: A Dynamic Speed Limit Control Algorithm based on Shock Wave Theory. *Proc., 11th International IEEE Conference on Intelligent Transportation Systems*, Beijing, China, 2008, pp. 827–832.
9. Carlson, R. C., I. Papamichail, and M. Papageorgiou. Local Feedback-based Mainstream Traffic Flow Control on Motorways Using Variable Speed Limits. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, No. 4, 2011, pp. 1261–1276.
10. Carlson, R. C., I. Papamichail, M. Papageorgiou, and A. Messmer. Optimal Motorway Traffic Flow Control Involving Variable Speed Limits and Ramp Metering. *Transportation Science*, Vol. 44, 2010, pp. 238–253.
11. Carlson, R. C., I. Papamichail, M. Papageorgiou, and A. Messmer. Optimal Mainstream Traffic Flow Control of Large-Scale Motorway Networks. *Transportation Research Part C*, Vol. 18, No. 2, 2010, pp. 193–212.

12. Chen, D., S. Ahn, and A. Hegyi. Speed Limit Control for Steady and Oscillatory Queues at Fixed Freeway Bottlenecks. Under review. 2014.
13. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves. I. Flood Movement in Long Rivers. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 229, 1955, pp. 281–316.
14. Richards, P. I. Shock Waves on the Highway. *Operations Research*, Vol. 4, 1956, pp. 42–51.
15. Bertini, R. L., and M. T. Leal. Empirical Study of Traffic Features at a Freeway Lane Drop. *Journal of Transportation Engineering*, Vol. 131, 2005, pp. 397–407.
16. Cassidy, M. J., and R. L. Bertini. Some Traffic Features at Freeway Bottlenecks. *Transportation Research Part B*, Vol. 33, No. 1, 1999, pp. 25–42.
17. Knoop, V. L., S. P. Hoogendoorn, and K. Adams. Incidents at Motorways: Capacity Reductions and the Effects of Incident Management. *European Journal of Transportation and Infrastructure Research*, Vol. 9, No. 4, 2009, pp. 363–379.
18. Laval, J. A., and C. F. Daganzo. Lane-changing in Traffic Streams. *Transportation Research Part B*, Vol. 40, No. 3, 2006, pp. 251–264.
19. Hegyi, A. *Model Predictive Control for Integrating Traffic Control Measures*. PhD dissertation. Delft University of Technology, Delft, The Netherlands, 2004.

TRAFFIC CONTROL

A Real-Time Signal Control Strategy for Mitigating the Impact of Bus Stops on Urban Arterials

CELESTE CHAVIS

Morgan State University

ELENI CHRISTOFA

University of Massachusetts, Amherst

Buses stopping at transit stops reduce the capacity of signalized intersections, which can lead to excessive delays for all users. In order to avoid such phenomena signal control strategies can be used. This paper presents a signal control strategy to mitigate the impact of bus stop operations on traffic operations along an undersaturated approach. The objective of the proposed strategy is to increase the green time for the bus stop approach during the cycle after the bus has left the stop in order to ensure that the residual queue that was created by the presence of the bus at the bus stop can fully dissipate within the following cycle. In addition, this strategy ensures that the cross-street approaches can clear any residuals queues caused by this strategy within a cycle after its implementation. Kinematic wave theory is used to track the formation and dissipation of queues and determine the red truncation (or equivalently green extension). The benefits achieved from the proposed strategy are illustrated through simulation tests at a single intersection for a variety of bus stop and bus operation characteristics. Average delay and average queue length for the bus stop and cross-street approaches are used to assess the performance of the system. The tests performed indicate that the signal control strategy can achieve substantial reductions in delay for the bus stop approaches without adversely affecting the cross-street operations and the overall intersection delay, when the demand at those cross streets is low.

INTRODUCTION

Efficient multimodal transportation systems are essential components for maintaining and improving the livability of our cities. However, the presence of multiple modes that differ in their dimensions and performance often complicates traffic operations and leads to underutilization of available capacity. Such an example is the existence of bus stops that are common in urban areas. If no bus bays exist, bus stops block lanes causing disruptions to the traffic stream, reducing the capacity of signalized intersections, and leading to excessive delays and potentially gridlock. Therefore, a comprehensive evaluation of the impact of bus stops on the capacity of signalized intersections and the development of signal control strategies to reduce this impact are critical for achieving efficient multimodal traffic operations and improving mobility for all users in urban networks.

The impact of bus stops on traffic and transit operations, and in particular on the capacity of signalized intersections has been extensively studied (1–8). However, several of the studies have not provided explicit formulas for estimating the impact (2, 5, 6), or have investigated the

impact in developing countries where bus stops are commonly located on the nonmotorized traffic lane and the behavior of drivers is substantially different than in the United States (1, 4, 7). Other studies have developed analytical formulas to estimate the impact as a function of the bus frequency and dwell time, the number of lanes, and the location of the subject bus stop with respect to the stop line (i.e., near side versus far side). However, some of these studies have not explicitly considered the impact of the bus stop's location (i.e., distance from the stop line) on the capacity of the signalized approach (3, 8). Only a recent analytical method has accounted for the location of the bus stop, the bus dwell time, and the bus frequency when determining the impact of bus stops on capacity (9).

While the impact of bus stops on the capacity of signalized intersections has been studied extensively, the literature on strategies that mitigate the impacts of bus stops on the capacity or delay of signalized intersections is very limited. A recent study by Gu et al. (10) investigated the impact of nearside bus stop location on the residual queue length and proposed a real-time bus holding strategy that ensures clearance of the queue within a signal cycle. However, the study used predicted arrival times instead of actual arrivals. The study was later extended (11) to estimate car and bus delays when the existence of bus stops affects their delays assuming stochastic bus arrivals and both nearside and far-side bus stops. However, the proposed formulas assumed that the car arrival flow is always less than the restricted capacity because of the presence of a bus at a bus stop, and researchers developed their formulas for certain distances of bus stops downstream and upstream of a signalized intersection.

Given the limitations of the literature, the objective of this study is to investigate the impact of bus stops on the capacity of signalized approaches as a function of the bus stop's location, bus arrival, and dwell times. The study investigates the impacts of those factors for cases that the car arrival flow is higher or lower than the restricted capacity when a bus is present at the bus stop during the green time interval, therefore, covering a variety of cases that can occur in reality. In addition, it suggests using information on the bus stop's impact on capacity to implement a real-time signal control strategy that ensures clearance of the residual queue within the cycle(s) following the detection of a bus's presence at a bus stop. The focus is on a well-timed signalized intersection for which all four approaches are undersaturated.

The rest of the paper is organized as follows: First, the authors describe the research approach that includes the methodology used to estimate the impact of an incident on traffic operations, in particular queue formation and dissipation as well as the proposed red truncation signal control strategy. Next, the authors present all different cases that can arise with regard to queue formation and dissipation patterns for a variety of incident locations, start times, and durations, as well as demand and signal timing characteristics. The required red truncation amounts are also calculated for a sample of cases. Next, results on the impact of the signal control strategy on the performance of the bus stop and cross-street approaches under a variety of demands, bus stop locations, and dwell times are presented. Finally, the authors comment on the applicability of the proposed strategy and suggest steps for extending the study.

RESEARCH APPROACH

The proposed research is based on tracking traffic conditions in the time–space domain while bus stops are both occupied and unoccupied by a bus. In order to do so kinematic wave theory (12, 13) is used. In particular it is assumed that traffic operations for an approach that contains a bus

stop can be described by a triangular fundamental diagram. By illustrating traffic conditions as states on a time–space diagram, it is possible to identify the formation and dissipation of queues and determine the impact of a bus stop on traffic conditions (e.g., queue length and delays).

Once traffic conditions are depicted on a time–space diagram, a signal control strategy is introduced. This strategy aims at increasing the green time for the bus stop approach over the next signal cycle in order to clear the additional queues created by the presence of the bus and avoid oversaturation of the subject approach. If the impact of the bus stop is small and the green time interval of the next cycle is sufficient to clear the residual queue, then no additional green time is provided for that approach over the next cycle. The proposed strategy can be implemented as an early green or green extension depending on the phase sequence within the cycle. Note that this strategy is based on the assumption that the cycle length remains constant for the cycle under consideration and the ones immediately following it. For illustrative purposes throughout this paper it is assumed that the green time interval follows the red in a cycle and therefore, the signal control strategy implemented corresponds to an early green (i.e., red truncation). In addition this study calculates the maximum red truncation that is allowed so that the cross street returns to undersaturated conditions within a cycle after the red truncation.

The proposed methodology assumes knowledge of the triangular fundamental diagram, a constant demand level for the subject approach, q_A , for the cycle under consideration and the ones immediately following it as well as the reduced capacity, q_I , caused by a bus dwelling at a bus stop; see Figure 1. It has also been assumed that the reduced capacity q_I is equivalent to the capacity of a traffic lane, which indicates that a bus dwelling at a stop would block one lane. In addition, the exact bus stop location as measured from the stop line, X , is known. Note that X is negative. It is also assumed that detection technologies such as automated vehicle location systems exist and can provide information on the bus arrival at the bus stop, T_o , and its departure from the stop, T_e , in real time. Note that no prediction of arrival or departure time is necessary

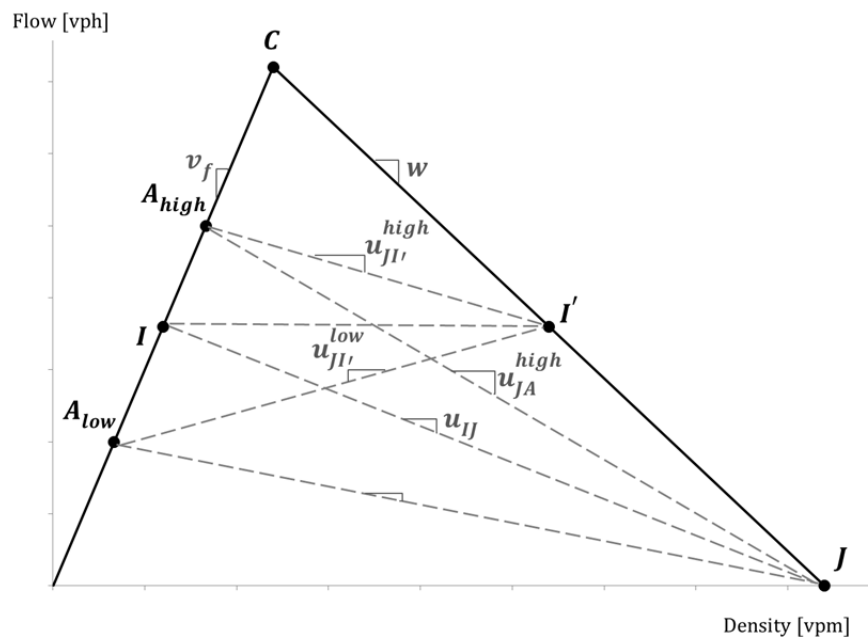


FIGURE 1 Fundamental diagram.

since the signal control strategy is always implemented in the cycle following the bus departure from the stop. It is further assumed that the bus dwell time does not exceed a cycle length, which is a reasonable assumption for most bus stops. Finally, it has been assumed that vehicle demand is available through sensing technologies such as loop detectors placed upstream of potential queue spillbacks, and the uncontrolled signal timing parameters of the pre-timed signal are known.

Red Truncation Estimation

In the absence of a bus at the stop the number of vehicles arrived and therefore, served in one cycle length, N , is equal to $q_A C/3,600$ where C is the cycle length in seconds and is the sum of the red time, R , and green time, G . However when a bus is present and oversaturation occurs, the number of vehicles, N_i , served can be determined by examining the flow rates at the stop line using the following equation:

$$N_i = (\tau_C q_C + \tau_I q_I + \tau_A q_A)/3,600 \quad (1)$$

where τ_C , τ_I , and τ_A is the total time that the flow rate is q_C , q_I , and q_A , respectively, during the cycle(s) that the bus is stopped at a bus stop. Given that the number of vehicles which were served by the intersection, N_i , is less than the number of vehicles that arrived, N_o , during the cycle(s) the incident is present, the amount that the red time interval is shortened is such that the number of vehicles that were unable to go through the intersection as a result of the bus stop-induced reduced capacity ($N_o - N_i$) plus the number of vehicles arriving in the next cycle (N) is equal to the number of vehicles that can be served during the initial green time for the bus approach, G , plus the additional green provided, $-D_R$. This can be expressed as follows:

$$D_R = \min \left\{ G - 3600 \frac{N_o - N_i + N}{q_C}, 0 \right\} \quad (2)$$

Where D_R is given in seconds and N_o and τ_C are based on whether or not the incident spans one or two cycles yielding

$$N_o = \begin{cases} N & \text{if } T_e \leq C \\ 2N & \text{otherwise} \end{cases} \quad (3)$$

$$\tau_C = \begin{cases} G - \tau_I - \tau_A & \text{if } T_e \leq C \\ 2G - \tau_I - \tau_A & \text{otherwise} \end{cases} \quad (4)$$

The equations for τ_I and τ_A are dependent on the cases defined by the position of the bus stop, the arrival and departure time of the bus from the stop, as well as the demand levels and signal timings, and they will be determined in the following section. Note that D_R is negative, because it corresponds to the change in the red time for the bus approach. If it becomes positive, then no truncation is needed because the initial green time for the approach G is sufficient to serve all vehicles despite the reduced capacity because of the presence of a bus at the stop.

Impact on the Cross-Street Traffic

The red truncation implemented for the bus stop approach, D_R , will result to a reduced green for the cross-street approaches. Assuming undersaturated traffic conditions for the cross streets, the maximum cross-street degree of saturation that will allow it to return to undersaturated conditions within a cycle after the implementation of the red truncation for the bus stop approach is determined. This maximum red truncation time, D_{Rmax} , can be calculated as

$$D_{Rmax} \geq \frac{2q_{Ax}C}{q_{Cx}} - 2G_x \quad (11)$$

where q_{Ax} is the arrival rate at the cross street and q_{Cx} is the saturation flow for the cross-street approach x , and G_x is the green time interval for cross-street approach x when no red truncation is implemented for the bus stop approach.

Identifying the Cases

Depending on the location of the bus stop, starting time, and dwell time of the bus, the capacity reduction can be categorized as falling into one of five cases. All other instances that do not fall into one of the five cases do not require red truncation (i.e., $D_R = 0$). In order to determine which case each bus stop induced capacity reduction incident is under and estimate its associated red truncation time, the following critical times have been identified as

$$T_1 = \frac{3600}{5280} \cdot \frac{X}{U_{JA}} \quad (5)$$

$$T_2 = R + \frac{3600}{5280} \cdot \frac{X}{w} \quad (6)$$

$$T_3 = \frac{q_c}{q_c - q_A} \cdot R + \frac{3600}{5280} \cdot \frac{X}{v_f} \quad (7)$$

$$T_4 = C + \frac{3600}{5280} \cdot \frac{X}{v_f} \quad (8)$$

$$T_5 = \frac{v_f T_o (q_A - q_I) + 3600/5280 X (q_c - q_A) + v_f q_c R}{(q_c - q_I) v_f} \quad (9)$$

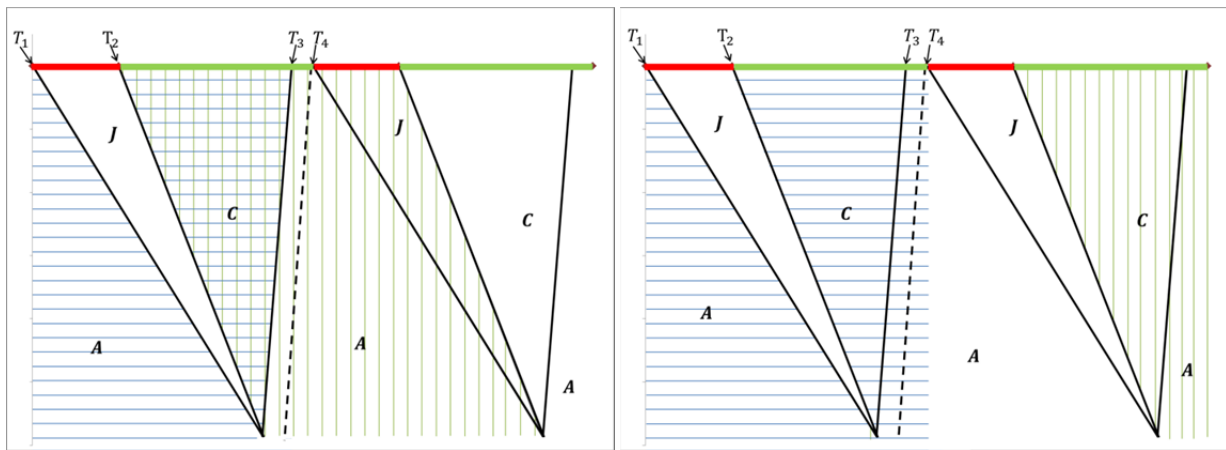
$$T_6 = \frac{T_e (q_I - q_C) + T_o (q_A - q_I)}{q_A - q_C} \quad (10)$$

where $T_1, T_2, T_3, T_4, T_5, T_6$ are given in seconds. These times are measured at the location of the dwelling bus, X .

Table 1 provides the constraints that are used to identify each of the cases. All constraints in each case must be satisfied, and each case is mutually exclusive. **Figure 2** shows the constraints for Cases 1 to 3 on a time–space diagram for undersaturated conditions as well as the locations of T_1 through T_4 . Note that for illustration purposes the yellow time intervals are not shown in the time–space diagrams of Figures 2 through 8 but are considered to be present at the end of each phase. The area with the horizontal lines represents the possible locations that T_o may be located and the vertical lines the locations where T_e may be located (subject to constraint

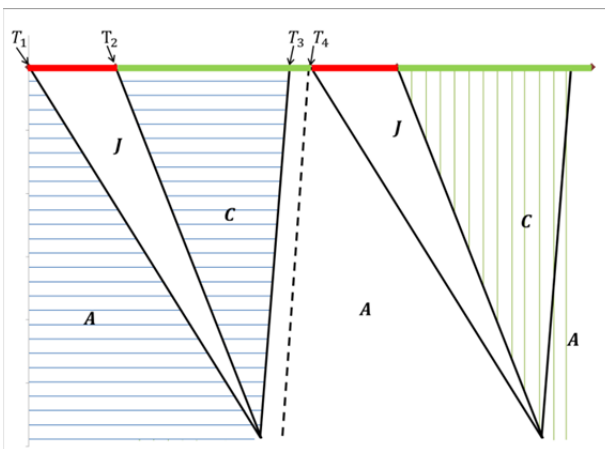
TABLE 1 Constraints for Each Case

Case	Constraints
1	$T_0 \leq T_3, T_2 \leq T_e \leq T_2 + C, T_5 > T_2$
2	$T_0 \leq C, T_e \geq T_2 + C, T'_5 > T_2$ if $q_A > q_I$ $T_0 \leq T_3, T_e \geq T_2 + C,$ otherwise
3	$T_3 < T_0 \leq C, T_2 + C \leq T_e \leq 2C, T_1 \leq T_2, q_A \leq q_I$
4	$T_5 \leq T_2, T_6 \geq T_4, T_0 < T_3, q_A > q_I$
5	$T'_5 \leq T_2, T_6 \geq T_4 + C, T_0 \geq T_3, q_A > q_I$

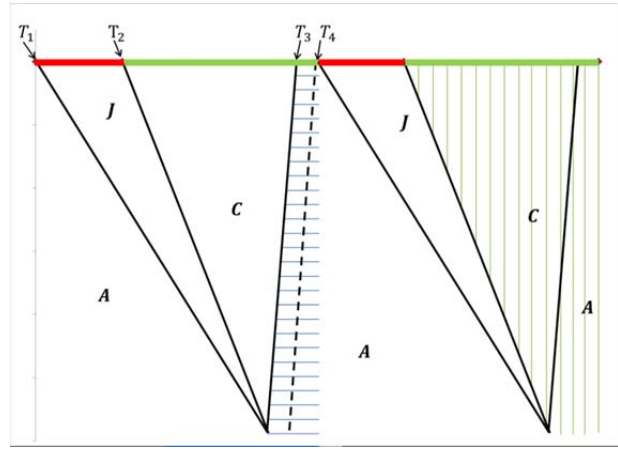


Case 1

Case 2



Case 2



Case 3



Possible locations



Possible locations

FIGURE 2 Illustration of constraints for Cases 1 through 3.

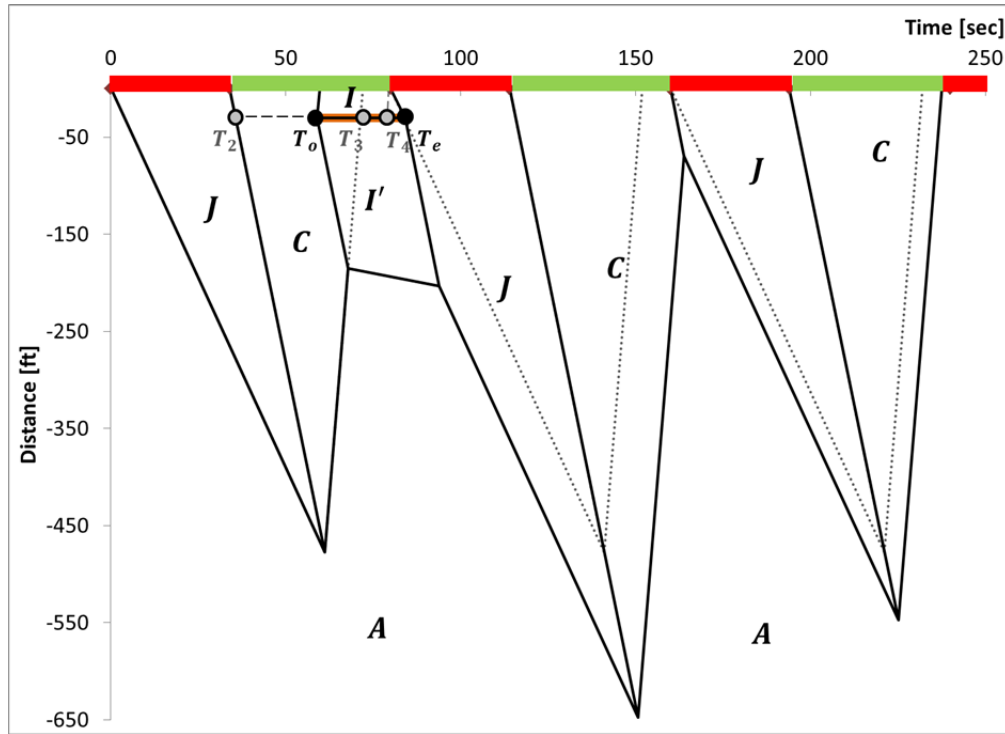


FIGURE 3 Case 1, time-space diagram (no control).

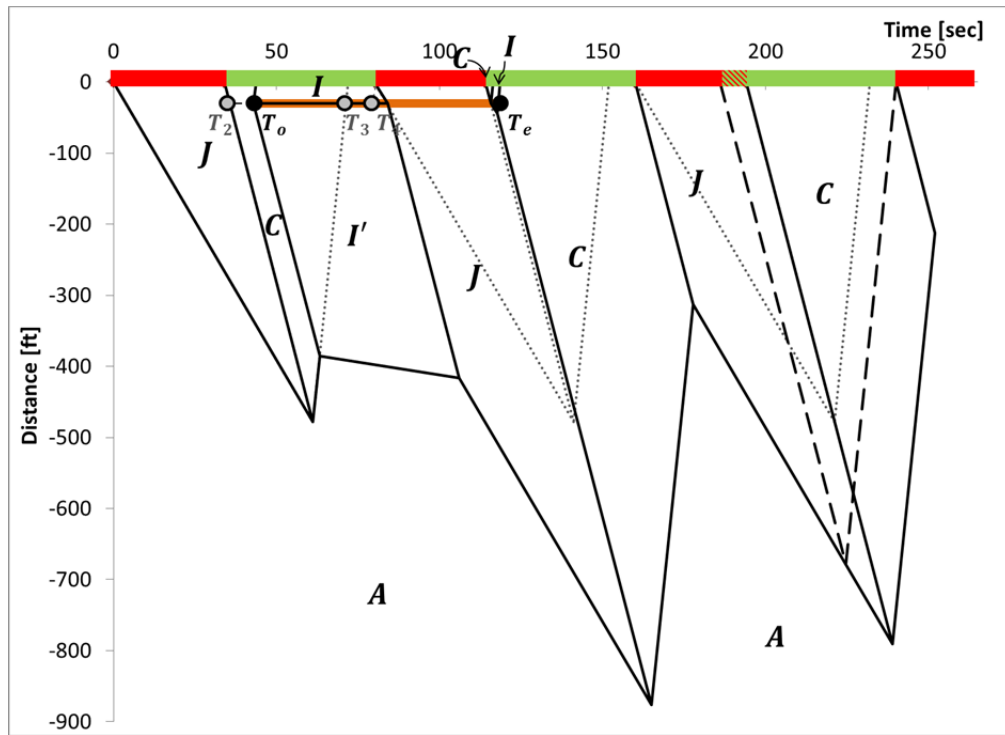


FIGURE 4 Case 2, time-space diagram (truncation needed).

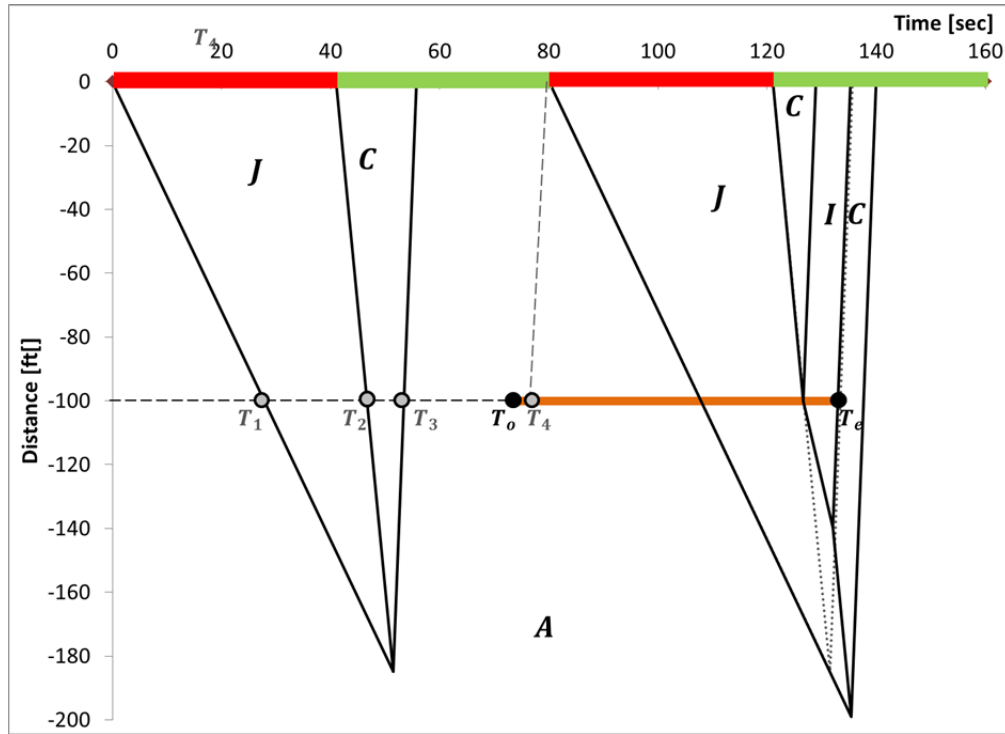


FIGURE 5 Case 3, time-space diagram (no control).

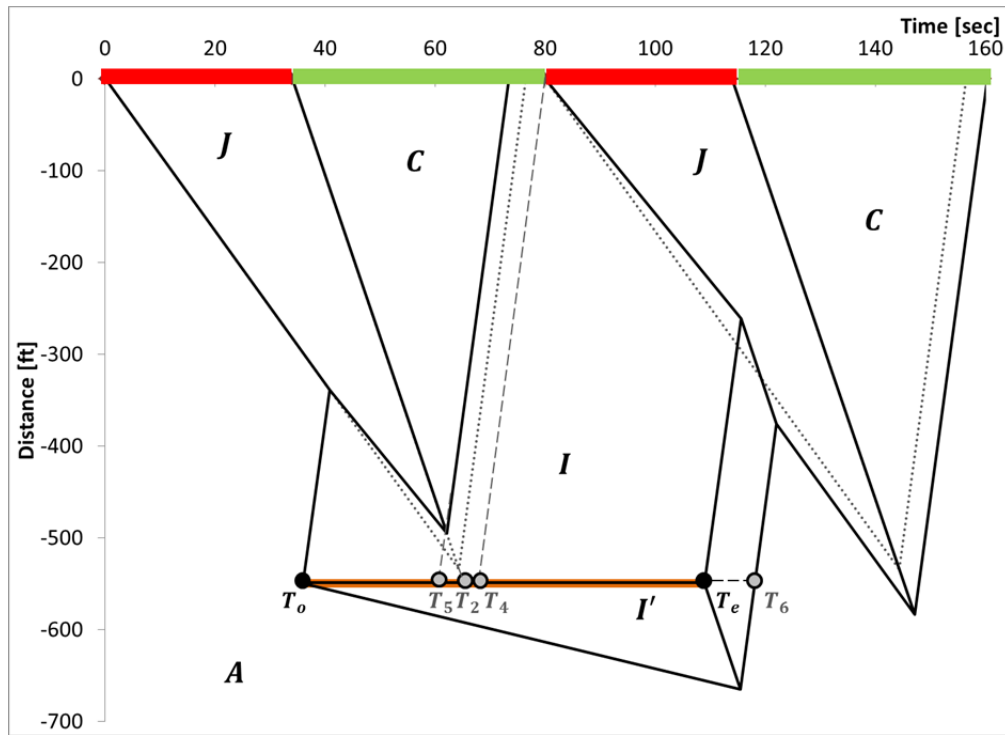


FIGURE 6 Case 4, time-space diagram (no control).

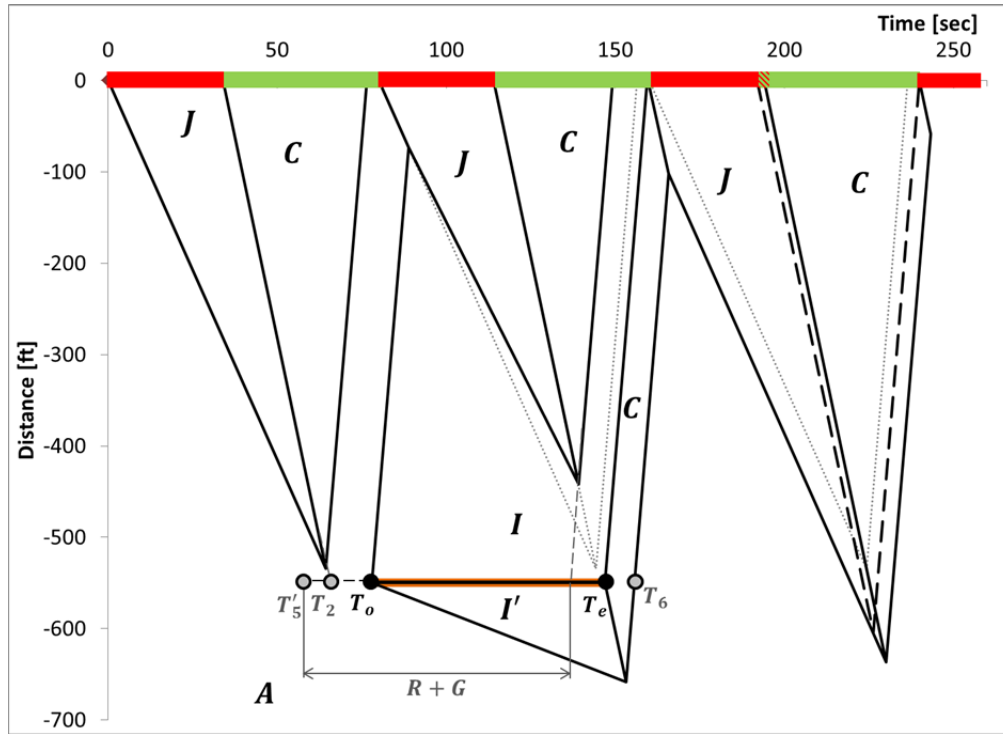


FIGURE 7 Case 5, time-space diagram (no control).

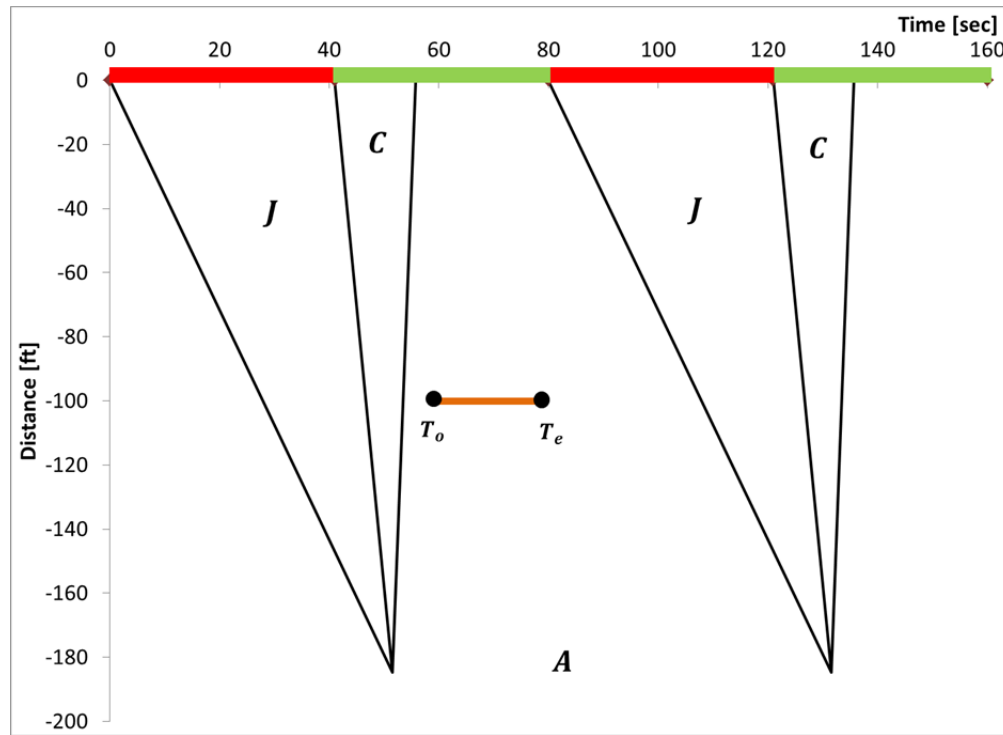


FIGURE 8 Case 6, no impact (no control).

$T_e - T_o \leq C$). It is not feasible for an incident to begin at a location in the jam state, since a bus cannot proceed to the bus stop if it is within the zero speed jam state. Cases 4 and 5 are not shown, as they represent special cases when the queue following the start of the incident is starved because of the presence of that incident when demand exceeds the capacity during the incident (i.e., $q_A > q_I$).

Once the case has been determined the red truncation time, D_R , can be calculated using the values in Table 2, which shows how long the intersection is discharging at rates as defined by States A and I . Figures 3 through 8 show example time–space diagrams for each case and depict the critical times defined above. Each figure shows in solid black lines the propagation and dissipation of queues if no signal control strategy is implemented and in dashed lines how the queue dissipates when the red truncation strategy is in place. The gray dotted lines represent the formation and dissipation of queues in the absence of a bus at the bus stop; see Figure 4. The bold line in yellow represents the presence of a stopped bus. For all cases shown it is assumed that traffic operations can be described by a fundamental diagram with the following characteristics: $q_c = 3,600$ veh/h, $q_i = 1,800$ veh/h, $w = -12$ mph, and $v_f = 30$ mph. As can be seen by the time–space diagrams in Figures 6 and 7, Case 5 is nearly a translation of Case 4. In order to determine the red truncation time in Case 5, three new variables must be defined: $T'_e = T_e - C$, $T'_o = T_o - C$, and T'_5 can be found by substituting T'_o for T_o in Equation 9.

The characteristics of bus stop operations for each sample case and the required red truncation as well as the resulting average vehicles delays for the bus stop approach and the cross street with and without control are shown in Table 3. When demand is low (Case 3, Figure 5; and Case 6, Figure 8) no truncation is needed. In fact, if the bus is not present at the bus stop while the signal is discharging, there is no increased delay as a result of the bus dwelling. In Cases 1 and 4, the presence of a bus at the bus stop increases delays; however, queues dissipate within a cycle without control. Cases 2 and 5 demonstrate scenarios that require red truncation. As expected the strategy improves the average and total delays along the bus stop approach and slightly increases delays along the cross street. Additional results are given in the following section.

APPLICATION

The evaluation of the proposed methodology and real-time signal control strategy has been performed with the use of simulation, in particular with the software AIMSUN through its application programming interface (API). API allows for implementing the proposed signal

TABLE 2 Time Spent in States A , τ_A , and I , τ_I

Case	τ_A	τ_I
1	0	$\min\{T_e, T_4\} - \max\{T_o, T_2\}$
2	$\min\{T_o, T_4\} - \min\{T_3, T_o\}$	$\max\{T_o, T_4\} - T_o + T_e - (T_2 + C)$
3	$T_4 - T_3$	$\min\{T_e, T_4 + C\} - (T_2 + C)$
4	0	$\min\{T_e, T_4\} - T_5$
5	$T_4 - T_3$	$\min\{T'_e, T_4\} - T'_5$

TABLE 3 Parameters and Average Vehicle Delays for Each Sample Case

	X (ft)	T_o (s)	T_e (s)	Demand Case	D_R	Bus Stop Approach Average Delay (s/veh)		Cross-Street Average Delay (s/veh)	
						No Control	Control	No Control	Control
Case 1	-30	1819	1843	$q_A > q_I$	0.00	16.49	—	31.46	—
Case 2	-30	43	118	$q_A > q_I$	-7.97	38.63	35.11	27.11	28.73
Case 3	-100	314	373	$q_A < q_I$	0.00	15.86	—	25.82	—
Case 4	-550	605	684	$q_A > q_I$	0.00	20.91	—	26.29	—
Case 5	-550	78	147	$q_A > q_I$	-2.06	18.15	17.44	27.11	27.11
Case 6	-100	619	639	$q_A < q_I$	0.00	11.10	—	16.47	—

control strategy in real time and evaluating its performance through a variety of performance measures such as delay for the subject bus stop approach, delay for the cross streets, and average queue length for all approaches.

Test Site

The test site used for the application of the proposed signal control strategy is the intersection of San Pablo and University Avenues in Berkeley, California; see Figure 9. The intersection operates with a four-phase signal with a cycle length of 80 s. Six bus routes with headways that vary between 10 and 30 min on each route, travel through the intersection and stop at six bus stops located at different distances from their corresponding stop lines.

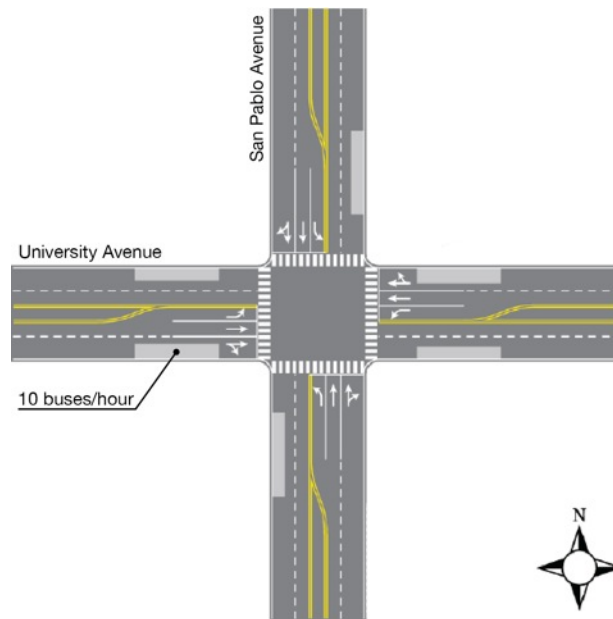


FIGURE 9 Test site: San Pablo and University Avenues, Berkeley, California.

The focus of this study is on the eastbound approach of University Avenue, which has a nearside bus stop where buses of three different bus lines stop with an overall frequency of 10 buses per hour. The demand levels of both approaches were set in such a way so that the test site represents traffic conditions at the intersection of a major with a minor roadway. This was essential since the proposed signal control strategy has been designed under the assumption of undersaturated conditions and performs best when the cross-street approach is not operating close to saturation. The car demand for the cross streets was set to 250 vph. Two scenarios of high and low demand were created for the bus stop approach (eastbound approach of University Avenue) and the opposite direction of University Avenue. The high demand was set to 2,000 vph and the low to 1,000 vph with turning ratios of 85% for through, 10% for right, and 5% for left for both demand scenarios. The same turning ratios were used for the cross-street demands.

For the low-demand scenario the green time for the phase that serves the University Avenue through movements was set to 37 s, and the one for the cross street through movements to 19 s. For the high-demand scenario, these green times were set equal to 44 s and 12 s, respectively. The other two phases that serve the left-turning vehicles had a constant green time of 5 s allocated to them for both scenarios. Therefore, red truncation was implemented only on the phase that serves the through cross-street movements. In addition, the lost time, which is assumed to be equal to the total yellow time, was kept constant and equal to 14 s, 4 s after each of the through phases and 3 s after each of the left-turning phases. It is assumed that cars will continue through the intersection for 2 s of the yellow interval, thus an extra 2 s was added the green time with remaining time in the cycle denoted as red time.

Based on this information, the degrees of saturation for the University Avenue through movements were 0.57 and 0.96 for the low- and high-demand scenarios and for the San Pablo approaches equal to 0.28 and 0.44 for the low- and high-demand scenarios. Bus stop locations of 30, 100, 200, and 500 ft from the stop line were tested. The average bus dwell time was set to 40 s with a standard deviation of 30 s. These ensured that several different cases of bus stop obstruction were captured.

Results

Several tests were performed with the help of the microsimulation software AIMSUN as described above, and the results were evaluated primarily through two performance measures: average vehicle delay and average queue length for both the bus stop approach and the cross streets (an average of the performance of the two cross-street directions is presented here). Ten replications were run for each scenario to account for the stochasticity in bus and vehicle arrivals as well as bus dwell times, and the average, as well as the 95% confidence intervals of those replications, are presented here.

Tests were performed both for a high bus stop approach demand, which exceeds the reduced capacity due to the bus dwelling, and for a low demand. The results for the low arrival rate along the bus stop approach show there is no need for red truncation at the cross street, since even with reduced capacity the intersection approach is still capable of serving all vehicles within one cycle, thus maintaining undersaturated conditions. It is possible that truncation is needed when the arrival rate is lower than the reduced capacity of intersection, as the presence of the bus may impede the discharge rate when a signal turns green. However, if the incident is sufficiently upstream and/or the degree of saturation low, the blockage caused by the bus dwelling at the stop does not greatly restrict the flow of arriving vehicles.

Figures 10 and 11 show the changes in average vehicle delay and average queue length for both the bus stop approach and the cross streets (presented as a weighted average of the two cross-street through approaches) as well as the 95% confidence intervals. The results indicate that the proposed signal control strategy can significantly reduce average vehicle delay for the bus stop approach for all bus stop location scenarios other than the one of 500 ft. As the distance of the bus stop from the stop line increases its impact on traffic operations close to the signalized intersection diminishes. As a result, even when no control is in place, the delays of vehicles on the bus stop approach are on average lower than the ones when the bus stop is located close to the intersection stop line. This verifies the findings of previous research efforts on the higher impact of nearside bus stops on the capacity and delays of vehicles at the signalized intersection. While the average vehicle delay for the cross streets increases, the amount by which it increases is very low, on the order of 3 to 5 s per vehicle. Taking into account that the cross street is less heavily traveled than the main street, the strategy results to an overall reduction of delay of about 6% for the case that the bus stop is located directly upstream of the intersection stop line.

Similar trends are observed for the average queue length at the bus stop and the cross-street approaches. The reduction of average queue length at the bus stop approach diminishes as the distance of the bus stop from the intersection stop line increases. At the same time the increase in average queue length observed at the cross-street approaches is minimal on the order of 0.2 vehicles. This happens because for the tests performed the cross streets were far from reaching saturation.

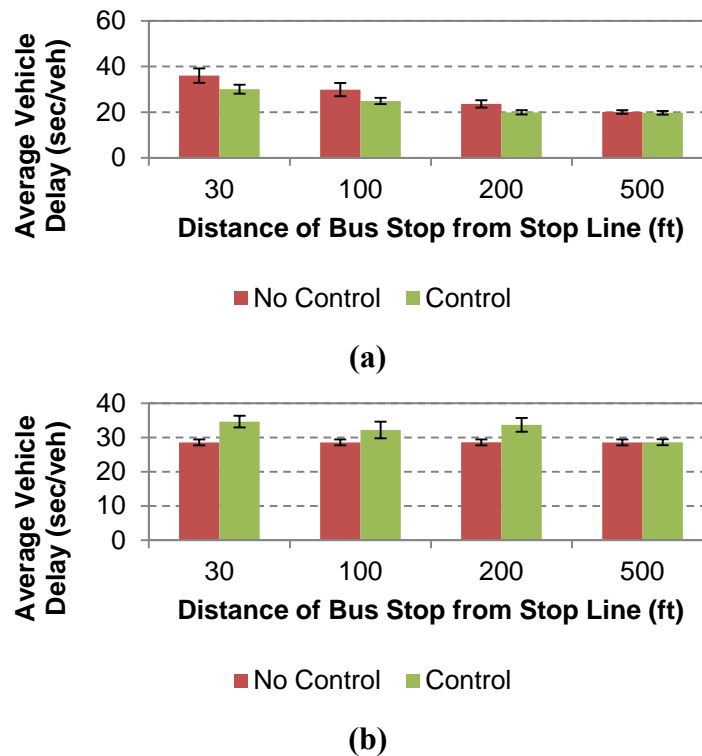
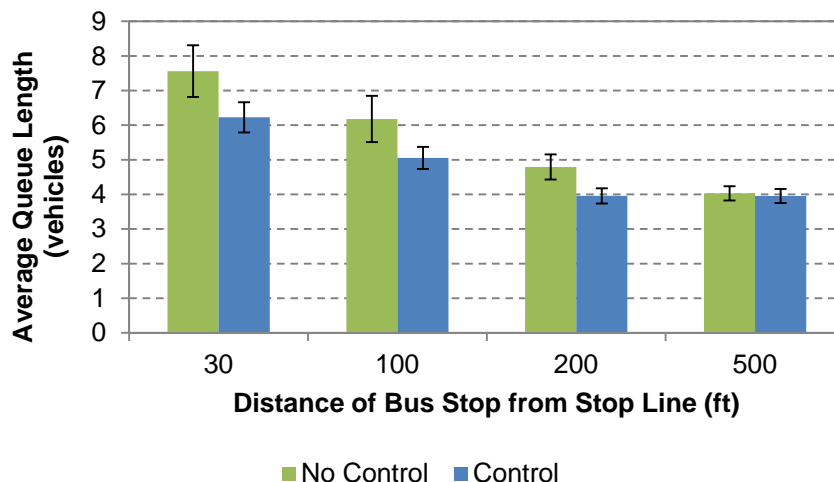
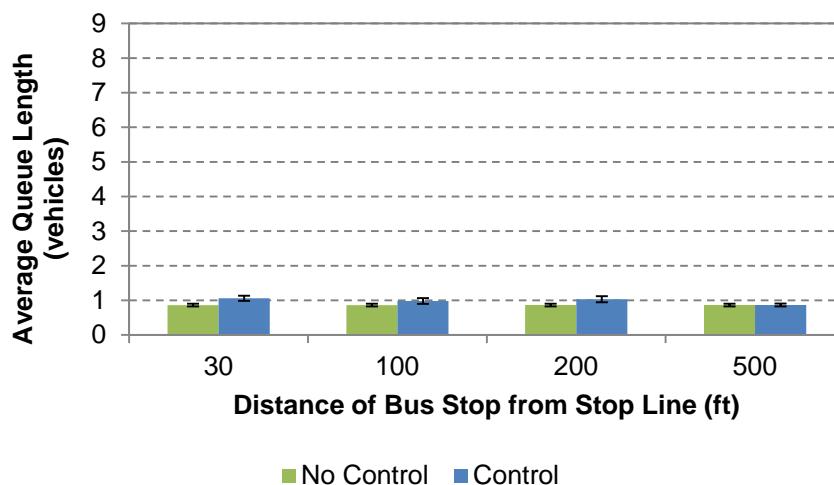


FIGURE 10 Average vehicle delay for various bus stop locations with and without control: (a) bus stop approach (eastbound University Avenue) and (b) cross-street approaches (average of northbound and southbound San Pablo Avenue).



(a)



(b)

FIGURE 11 Average queue for various bus stop locations with and without control: (a) bus stop approach (eastbound University Avenue) and (b) cross-street approaches (average of northbound and southbound San Pablo Avenue).

CONCLUSIONS

This study presents the development of a real-time signal control strategy that uses information on the location of a bus stop, as well as the bus dwell time, traffic demand levels, and signal timings. The proposed bus stop mitigation strategy uses kinematic wave theory to track the formation and dissipation of queues and estimate the amount of green that needs to be added to the subject approach so that residual queues are cleared within one cycle.

Several tests were performed for a variety of bus stop locations, dwell times, and for two levels of bus stop approach demands through microsimulation. The outcomes of the tests indicate that rarely if ever there is a need for implementing the real-time signal control strategy when the capacity of the approach when a bus is present is higher than the demand of the incoming traffic

to the subject approach. On the other hand, when the arrival demand at the bus stop approach exceeds the lower capacity caused by a bus dwelling at the bus stop, the proposed mitigation strategy can achieve average delay and average queue length reductions of up to 17% for the bus stop approach. The 17% reduction in average delay corresponds to about 6 s per vehicle average delay savings. At the same time, the mitigation strategy only slightly increases average delay for the cross street on the order of 3 to 5 s per vehicle. When the cross street is less heavily traveled than the bus stop approach, net benefits from such a control strategy are achieved, and the overall delay at the intersection decreases. The mitigation strategy is most beneficial when the bus stop is located very close to the stop line of the signalized approach. Note that the proposed strategy is applicable for cases where all intersection approaches are undersaturated and is most beneficial when implemented at intersections that the bus stop approach has a much higher demand than the cross street.

The benefit of the proposed real-time signal control strategy is that it can be implemented for any bus stop location, dwell time, and can be used when the vehicle demand is higher than the reduction in capacity induced by a bus stopping at bus stop and when it is lower. Therefore, in addition to investigating and mitigating the impact of bus stops on capacity and delay, the equations presented here and the mitigation strategy are applicable to any type of incident that can occur in signalized arterial networks, such as freight deliveries, accidents, and more. So, under the assumption that the characteristics of the incident are known in real time (after the incident had been removed), the proposed strategy can be implemented to mitigate the impact of that incident on traffic of that approach. Overall the proposed strategy can be used for real-time mitigation of bus stop or incident-induced reductions of capacity to improve traffic and transit operations in urban signalized arterials.

Future steps include improving the mitigation strategy so it can handle bus stop events that occur within the same cycle or consecutive cycles and cases when a bus is stopping for longer than one cycle length, and extending the focus of the strategy to include cases where vehicle demand varies from cycle to cycle.

ACKNOWLEDGMENT

This research was sponsored by the University of Massachusetts STEM Diversity Faculty Exchange Program.

REFERENCES

1. Coeymans, J. E., and J. C. Herrera. Estimating Values for Traffic Parameters in Turning Lanes. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1852, Transportation Research Board, Washington, D.C., 2003, pp. 47–54.
2. Tang, T. Q., Y. Li, and H. J. Huang. The Effects of Bus Stop on Traffic Flow. *International Journal of Modern Physics C*, 20(06), 2009, pp. 941–952.
3. Teply, S., D. I. Allingham, D. B. Richardson, and B. W. Stephenson. *Canadian Capacity Guide for Signalized Intersections*. 2008.
4. Yang, X., Z. Gao, X. Zhao, and B. Si. Road Capacity at Bus Stops with Mixed Traffic Flow in China. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2111, Transportation Research Board, Washington, D.C., 2009, pp. 18–23.

5. Zhao, X. M., Z. Y. Gao, and B. Jia. The Capacity Drop Caused by the Combined Effect of the Intersection and the Bus Stop in a CA Model. *Physica A: Statistical Mechanics and its Applications*, 385(2), 2007, pp. 645–658.
6. Zhao, X. M., Z. Y. Gao, and K. P. Li. The Capacity of Two Neighbour Intersections Considering the Influence of the Bus Stop. *Physica A: Statistical Mechanics and its Applications*, 387(18), 2008, pp. 4649–4656.
7. Zhao, X. M., B. Jia, Z. Y. Gao, and R. Jiang. Traffic Interactions Between Motorized Vehicles and Nonmotorized Vehicles near a Bus Stop. *Journal of Transportation Engineering*, 135(11), 2009, pp. 894–906.
8. *Highway Capacity Manual 2010*. Transportation Research Board of the National Academies, Washington, D.C., 2010.
9. Li, M., H. Chen, and Y. Zhang. Capacity Analysis of Road Segment near Bus Stop Affected by Signalized Intersection. In *CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient*, ASCE, 2012, pp. 683–694.
10. Gu, W., M. J. Cassidy, V. V. Gayah, and Y. Ouyang. Mitigating Negative Impacts of Near-side Bus Stops on Cars. *Transportation Research Part B*, 47, 2013, pp. 42–56.
11. Gu, W., V. V. Gayah, M. J. Cassidy, and N. Saade. *On the Impacts of Bus Stops near Signalized Intersections: Models of Car and Bus Delays*. Working Paper UCB-ITS-VWP-2014-1. UC Berkeley Center for Future Urban Transport, A Volvo Center of Excellence. 2014.
12. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 229(1178), 1955, pp. 317–345.
13. Richards, P. I. Shock Waves on the Highway. *Operations Research*, 4(1), 1956, pp. 42–51.

Freeway Traffic Analysis

FREEWAY TRAFFIC ANALYSIS

Influential Subspaces of Connected Vehicles in Highway Traffic**KSHITIJ JERATH****VIKASH V. GAYAH****SEAN N. BRENNAN***The Pennsylvania State University*

This work introduces the novel concept of an influential subspace, with focus on its application to highway traffic containing connected vehicles. In this context, an influential subspace of a connected vehicle is defined as the region of a highway where the connected vehicle has the ability to positively influence the macrostate (i.e., the traffic jam), so as to dissipate it within a specified time interval. Analytical expressions for the influential subspace are derived using the Lighthill-Whitham-Richards (LWR) theory of traffic flow. Included results describe the span of the influential subspace for specific traffic flow conditions and prespecified driving algorithms of the connected vehicles.

INTRODUCTION

In recent years, there have been significant developments in the ability to inform drivers about nearby traffic conditions, which often leads to the questions: Can an individual driver use such information to affect traffic flow? And which drivers in a traffic network have the most influence on traffic flow (i.e., where and to whom should the information be delivered)? This work introduces the concept of an influential subspace of a connected vehicle (I), which is defined as the region of a highway where the connected vehicle has the ability to drive the macroscopic state of traffic flow to a desired state within a prespecified time. This concept is applicable to several physical, biological, and engineered systems, and a general formulation will be presented in future publications. In this paper, analysis of the influential subspace is conducted specifically for a connected vehicle entering a self-organized traffic jam, using basic postulates of the LWR model of traffic flow (2, 3).

To better understand the concept, consider the following example with reference to **Figure 1**. Given a single-lane highway segment where no passing is allowed, assume that a spontaneous traffic jam has formed on one section so that the macroscopic state (or simply macrostate) of traffic flow in that region is the jammed state J . Next assume that the desired macrostate is free flow (state A), with known flow and density, that currently exists upstream of the traffic jam. Vehicles in this state are assumed to travel at the maximum permissible velocity [i.e., the free-flow velocity (v_f)], and cannot travel any faster. Now, as a thought experiment, consider the impact that a connected vehicle receiving information on downstream traffic conditions could have on the jammed state, for each of the four regions outlined in **Figure 1**.

In **Region 1** of **Figure 1**, a connected vehicle is sufficiently far from the jammed state so that its actions (such as slowing down to avoid the jam) have no positive effect on the jammed state—the jam would have dissipated by the time the connected vehicle of **Region 1** moves downstream. In **Region 2**, a connected vehicle could slow down to avoid the traffic jam, and this action could

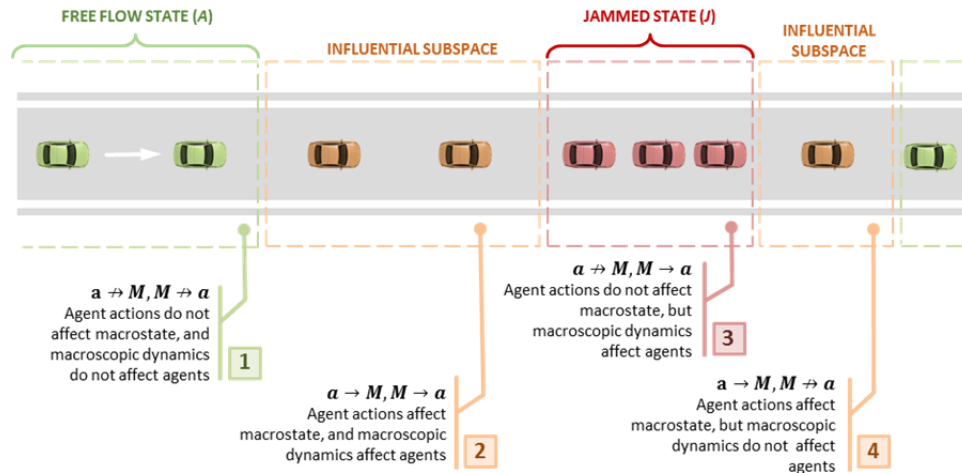


FIGURE 1 Thought experiment for understanding the concept of influential subspaces of connected vehicles in highway traffic. White arrow indicates direction of travel.

result in fewer vehicles entering the jammed state. As a result, Region 2 represents the influential subspace. However, as a connected vehicle moves closer to the jammed state J , its influence decreases, and in Region 3, the connected vehicle cannot escape the jam by slowing down and its actions have no positive influence on the jammed state. Finally, in Region 4, a vehicle may decide to exit the jam slower rather than at the free-flow velocity, resulting in a negative influence and a persistent jammed state. The outcomes of this thought experiment are validated using the LWR model in later sections.

As connected vehicles technology becomes sufficiently advanced and begins to enter the mainstream, it is imperative that the research community helps fully realize its potential and efficacy. Prior work on connected vehicles has primarily focused on communication protocols and vehicular network topologies. While this research is important, it produces few research insights into the potential impact of connected vehicles on traffic flow. Recent work on the impact of mixed traffic on self-organized jams (4), effect of individual driving strategies on traffic flow (5), cooperative adaptive cruise control (6), and cooperative highway driving (7) have all briefly touched on various aspects of how individuals affect macroscopic traffic flow dynamics. However, these research efforts do not address the traffic system from the perspective of influential subspaces of connected vehicles. The following section presents the framework within which the concept of influential subspaces will be introduced.

PROBLEM SETUP

The problem is setup as a single-lane highway where no passing is allowed. Representative values of traffic flow parameters such as maximum flow ($q_{\max} = 1,800$ veh/h), jam density ($k_j = 110$ veh/km), and free-flow velocity ($v_f = 90$ km/h) are used, assuming a triangular relationship between flow and density. The analysis uses standard results of the LWR model by drawing time–space diagrams to identify the time taken for the traffic flow to reach a desired macrostate (i.e., one in which the traffic system is operating in a free-flow state).

To keep the analysis simple, only two connected vehicles are considered in the presented

work. At time $t = 0$, the first connected vehicle ($CV1$) enters the jam and sends an alert signal indicating a jammed state to the connected vehicle upstream, which receives the signal instantaneously. The reception of the alert signal from $CV1$ causes an event-triggered control action in $CV2$, which slows down to a predetermined speed v_s as selected by the driver or dictated by an inbuilt cruise control algorithm. When $CV1$ exits the traffic jam at time $t = t_{\text{EXIT}}$, it sends another alert signal upstream. This alert results in a second event-triggered control action in $CV2$ due to which it speeds up to free-flow velocity v_f . Depending on the location of the second connected vehicle $CV2$, its control policy (i.e., the combined event-triggered actions of slowing down and speeding up) may or may not have an effect on the macrostate. The next section discusses several explanatory cases similar to the ones described in Figure 1 that make the problem setup clearer.

INFLUENTIAL SUBSPACES OF CONNECTED VEHICLES

For the following example, the traffic system is assumed to be operating at traffic state A given by $q_A = 900$ veh/h and $k_A = 10$ veh/km. Without loss of generality, it may be assumed that the first connected vehicle $CV1$ enters the spontaneous traffic jam and sends the alert signal at time $t = 0$. Upon receiving the signal, the second connected vehicle $CV2$ is assumed to slow down to a predetermined speed v_s 10 km/h in order to avoid the traffic jam. This results in a slow-moving state S given by $q_S \approx 733$ veh/h and $k_S = 73$ veh/km.

Interpretation of Time–Space Diagrams

With reference to Figures 1 and 2, parts a and b correspond to Region 3 in Figure 1. In these cases, the actions of the vehicle $CV2$ have no positive effect on the time it takes to return to the desired macrostate A . In both cases, the jammed state J dissipates at time t_j , independent of the presence of connected vehicles in the traffic stream. Figure 2c corresponds to Region 2 in Figure 1, where the actions of vehicle $CV2$ cause the traffic system to reach the desired macrostate A faster. Specifically, the slow-moving state S vanishes at time t_s , whereas the jammed state vanishes at time $t_j < t_s$. Thus, there is a net reduction in the time taken for the traffic flow to return to the desired macrostate A . Finally, Figure 2d corresponds to Region 1 of Figure 1, where the actions of vehicle $CV2$ have no positive impact on the time taken to return to macrostate A , since the jammed state J dissipates of its own accord.

Analytical Solution of Influential Subspaces

Mathematically, the time taken for the traffic system to reach the desired macrostate A is given by

$$t_A = \max\{t_J, t_S\} \quad (1)$$

where t_J denotes the time taken for the jammed state J to dissipate, and t_S represents the time taken for the slow-moving traffic state S to vanish. In other words, the time taken to reach the desired macrostate A is governed by which of state J or S persists for a longer period of time. The mathematical expressions for t_J and t_S can be calculated from geometric considerations of Figure 2.

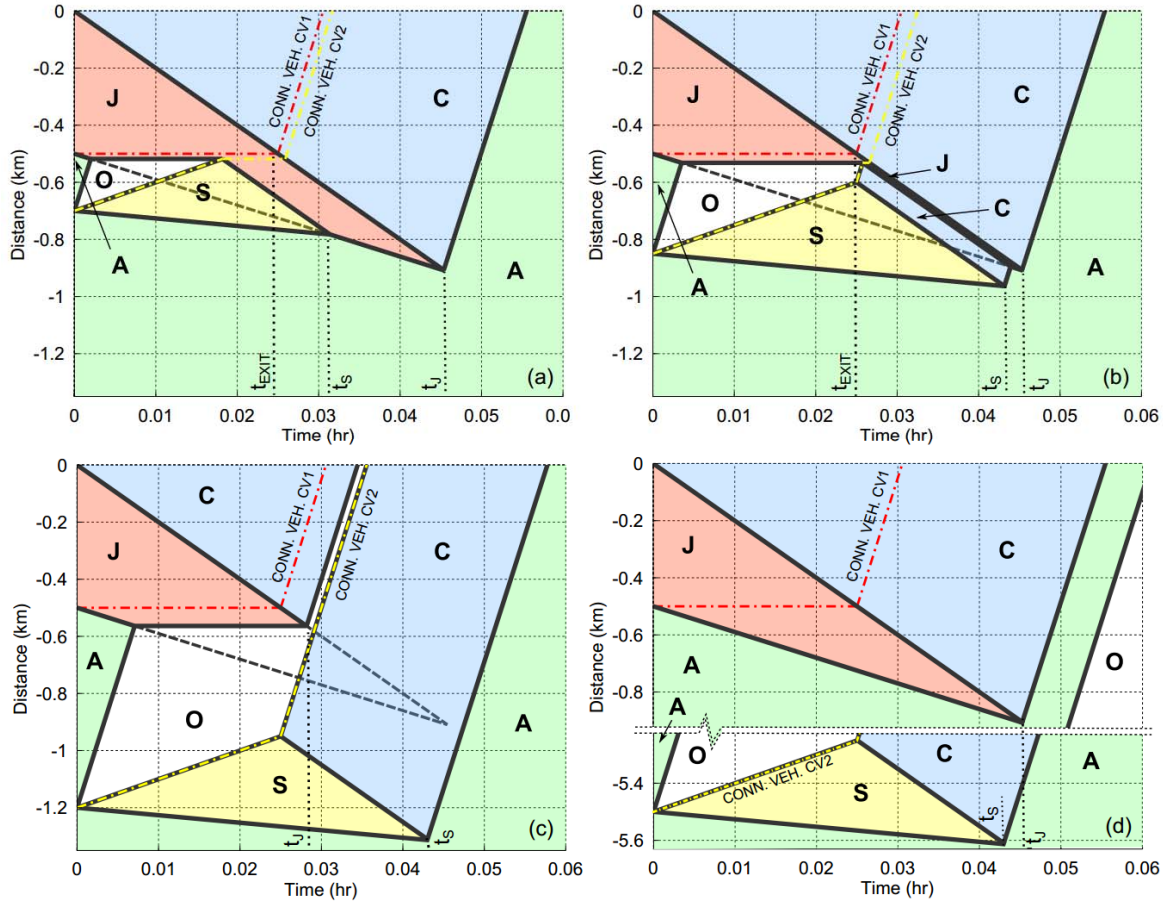


FIGURE 2 Time–space diagram when distance between the connected vehicles $CV1$ and $CV2$ is (a) 200 m, (b) 350 m, (c) 700 m, and (d) 5,000 m, for $v_s = 10$ km/h. In cases (a) and (b), the actions of $CV2$ have no positive impact on time taken for traffic to return to macrostate A . In case (c), the slowing down by $CV2$ causes a more rapid return to macrostate A (e.g., jam-free traffic flow); $CV2$ is in its influential subspace. In case (d), the vehicle $CV2$ has no positive impact on the macrostate; the jam has already dissipated. A dashed line indicates jam evolution without connected vehicles. Dash-dotted lines are vehicle trajectories of connected vehicles.

Expression for Dissipation Time t_j of Jammed State J

Specifically, first consider the evaluation of t_j with reference to **Figure 3** (or **Figure 2c**). In this scenario, the time taken for the jammed state J to dissipate is a function of the original queue length x_q at time $t = 0$, the distance between the connected vehicles x_d at time $t = 0$, and the traffic state A that exists upstream of the jammed state J . The expression for t_j in **Figure 3** is given by **Equation 2** as follows:

$$t_j = \frac{x_q + x_n}{w} \quad (2)$$

where x_n is the length of the roadway occupied by new vehicles entering the jammed state J after time $t = 0$, and w is the backward wave speed obtained from the triangular fundamental diagram. The quantity x_n is determined by assuming that the number of vehicles is conserved on the roadway. Specifically, under this assumption, the number of vehicles between the two connected vehicles $CV1$ and $CV2$ can be calculated to be

$$\text{Number of vehicles between } CV1 \text{ and } CV2 = x_d k_A = x_n k_J \Rightarrow x_n = \frac{x_d k_A}{k_J} \tag{3}$$

where x_d is the distance between the connected vehicles $CV1$ and $CV2$, and k_A and k_J represent the densities of traffic flow in states A and J , respectively. Consequently, the expression in Equation 2 can be expanded to yield

$$t_J = \frac{x_q + x_d k_A / k_J}{w} \tag{4}$$

However, this expression is correct only for a specific region of the roadway. The analytical expressions demarcating this specific region can be found by a careful analysis of Figure 3. Note that the expression for t_J in Equation 4 becomes valid in situations similar to Figure 3, when the second connected vehicle just manages to avoid the jammed state J , and stays valid till situations similar to Figure 2d, when the last vehicle ahead of the vehicle $CV2$ just manages to avoid the

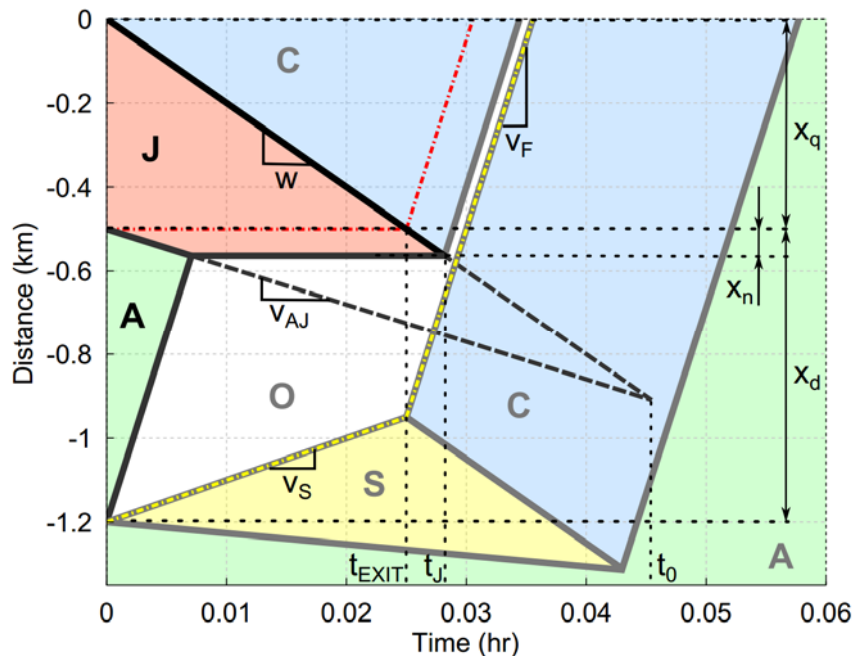


FIGURE 3 Evaluation of t_J using space–time diagram. Only relevant quantities needed for deriving analytical solution are labeled.

jammed state J . To evaluate the lower spatial limit (i.e., in the case when the second connected vehicle just manages to avoid the jammed state J), Equation 4 becomes valid if

$$x_d - x_n \geq v_s t_{EXIT} + v_f (t_j - t_{EXIT}) \quad (5)$$

where v_s represents the speed that the second connected vehicle $CV2$ slows down to, v_f represents the free-flow velocity, and $t_{EXIT} (= x_q/w)$ represents the time at which the first connected vehicle $CV1$ exits the jammed state J . The expression in Equation 5 may be simplified to be written as:

$$x_d - x_n \geq v_s \frac{x_q}{w} + v_f \left(\frac{x_q + x_d k_A / k_J}{w} - \frac{x_q}{w} \right) \quad (6)$$

$$\text{or } x_d - x_n \geq v_s \frac{x_q}{w} + \frac{v_f}{w} \left(\frac{x_d k_A}{k_J} \right) \quad (7)$$

$$\text{or } x_d - x_d \frac{k_A}{k_J} - x_d \left(\frac{v_f k_A}{w k_J} \right) \geq v_s \frac{x_q}{w} \quad (8)$$

$$\text{or } x_d \geq \left\{ 1 - \left(1 + \frac{v_f}{w} \right) \frac{k_A}{k_J} \right\}^{-1} \left(v_s \frac{x_q}{w} \right) \quad (9)$$

The upper spatial limit for the validity of Equation 4 is evaluated in the scenario when the second connected vehicle $CV2$ is sufficiently upstream so that last vehicle just ahead of $CV2$ reaches the jammed state at time t_0 , (i.e., when the jam is just about to dissipate of its own accord). Thus, the upper spatial limit is given simply by

$$x_q + x_n \leq w t_0 \quad (10)$$

$$\text{or } \frac{k_A}{k_J} x_d \leq w t_0 - x_q \quad (11)$$

$$\text{or } x_d \leq \frac{k_J}{k_A} (w t_0 - x_q) \quad (12)$$

On the other hand, in Figure 2 parts a , b , and d , the expression for t_j is obtained quite simply from the original jam dissipation time t_0 evaluated in the absence of any connected vehicles. The jam evolution trajectory is indicated using dashed lines in Figures 2 and 3. In these scenarios, the jam dissipation time $t_j = t_0$ and is found as follows:

$$\text{Distance traveled} = w t_0 = x_q + v_{AJ} t_0 \Rightarrow t_0 = \frac{x_q}{w - v_{AJ}} \quad (13)$$

where v_{AJ} represents the interface speed between traffic states A and J . Consequently, the expression for time taken for dissipation of the jammed state J is given by combining the expressions in Equations 4, 9, 12, and 13 to yield

$$t_J = \begin{cases} \frac{1}{w} \left\{ x_q + x_d \left(\frac{k_A}{k_J} \right) \right\}, & \text{if } \left\{ 1 - \left(1 + \frac{v_f}{w} \right) \frac{k_A}{k_J} \right\}^{-1} \left(v_s \frac{x_q}{w} \right) \leq x_d \leq \frac{k_J}{k_A} (w t_0 - x_q) \\ \frac{x_q}{w - v_{AJ}}, & \text{else} \end{cases} \quad (14)$$

Expression for Dissipation Time t_S of Slow-Moving State S

Similar geometric arguments can be used to determine the expression for the time taken for the slow-moving traffic state S to dissipate. Specifically, consider Figure 4 (or Figure 2a) in order to ascertain the analytical expressions.

If the second connected vehicle $CV2$ is too close to the first one, as depicted in Figure 4 (or Figure 2a), it enters the jam and the dissipation time for state S is governed by this distance. In alternative scenarios, when the vehicle $CV2$ is further upstream, the dissipation time is constant, as evinced by Figure 2 parts b , c , and d . In Figure 4, the dissipation time of the slow-moving state can be evaluated by geometric calculations as follows:

$$\text{Distance} = v_s t_{HIT} - w(t_S - t_{HIT}) = v_{AS} t_S \Rightarrow t_S = \left(\frac{v_s + w}{v_{AS} + w} \right) t_{HIT} \quad (15)$$

Where t_{HIT} is the time at which the vehicle $CV2$ first enters the jammed state J , and v_{AS} is the interface speed between the states A and S . The expression for t_{HIT} can be found using geometric considerations to be

$$t_{HIT} = \frac{x_d - x_n}{v_s} = \frac{x_d}{v_s} \left(1 - \frac{k_A}{k_J} \right) \quad (16)$$

so that the dissipation time of state S when $CV2$ is close to the jam is given by

$$t_S = \left(\frac{v_s + w}{v_{AS} + w} \right) \left(1 - \frac{k_A}{k_J} \right) \frac{x_d}{v_s} \quad (17)$$

On the other hand, in Figure 2 parts b , c , and d , where the vehicle $CV2$ is further upstream, the dissipation time for the state S can be calculated similarly as follows:

$$\text{Distance} = v_s t_{EXIT} - w(t_S - t_{EXIT}) = v_{AS} t_S \Rightarrow t_S = \left(\frac{v_s + w}{v_{AS} + w} \right) t_{EXIT} \quad (18)$$

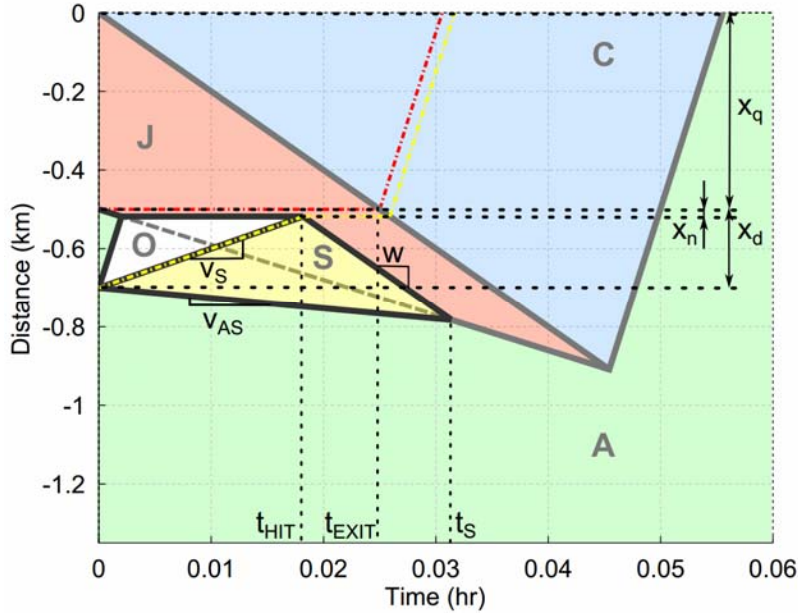


FIGURE 4 Evaluation of t_s using space–time diagram. Only relevant quantities needed for deriving analytical solution are labeled.

where t_{EXIT} is the time at which the first connected vehicle exits the jammed state J , and which can be found using geometric considerations to be

$$t_{EXIT} = \frac{x_q}{w} \quad (19)$$

so that the dissipation time of state S when $CV2$ is close to the jam is given by

$$t_s = \left(\frac{v_S + w}{v_{AS} + w} \right) \frac{x_q}{w} \quad (20)$$

Consequently, by observing the nature of t_s across the various parts of Figure 2, it is realized that the expression for the dissipation time for the slow-moving state S is

$$t_s = \min \left\{ \left(\frac{v_S + w}{v_{AS} + w} \right) \left(1 - \frac{k_A}{k_J} \right) \frac{x_d}{v_S}, \left(\frac{v_S + w}{v_{AS} + w} \right) \frac{x_q}{w} \right\} \quad (21)$$

To recapitulate the major result of this work, the time taken for the traffic system to reach the desired macrostate A , is given by

$$t_A = \max\{t_J, t_s\} \quad (22)$$

where the expressions for t_J and t_s are provided in Equations 14 and 21, respectively.

RESULTS

An influential subspace is defined for a specific agent or vehicle in a multiagent system. The influential subspace is defined by the ability of the specific agent or vehicle to drive the system to a desired macrostate (A) within a predetermined time (t_{DES}). In this example, the time taken for the traffic system to reach the macrostate A is calculated for varying distances between the connected vehicles $CV1$ and $CV2$. If the goal is to reach the macrostate A within, say $t_{DES} = 160$ s, then the influential subspace for $CV2$ is situated between 0.5 and 4.3 km from the vehicle $CV1$, as indicated in Figure 5. On the other hand, if the goal is to reach the macrostate A within, say $t_{DES} = 100$ s, then it can be said that the influential subspace is empty, or it does not exist.

Knowledge of the influential subspace is a critical element for the efficient implementation of connected vehicles technology. Implementation of connected vehicles technology will have to deal with, among other things, issues such as bandwidth limitations and packet transmission ranges. Consequently, knowledge of the influential subspace can help ensure that bandwidth is not wasted by transmitting packets to vehicles that are not in their influential subspaces. Additionally, the same knowledge can help optimally route packets to vehicles within the influential subspaces and reduce power requirements for transmission equipment. Further, the concept of influential subspaces has significant potential applications in other areas such as cooperative adaptive cruise control, where formation, merging, and splitting of platoons can benefit from the use of this novel concept.

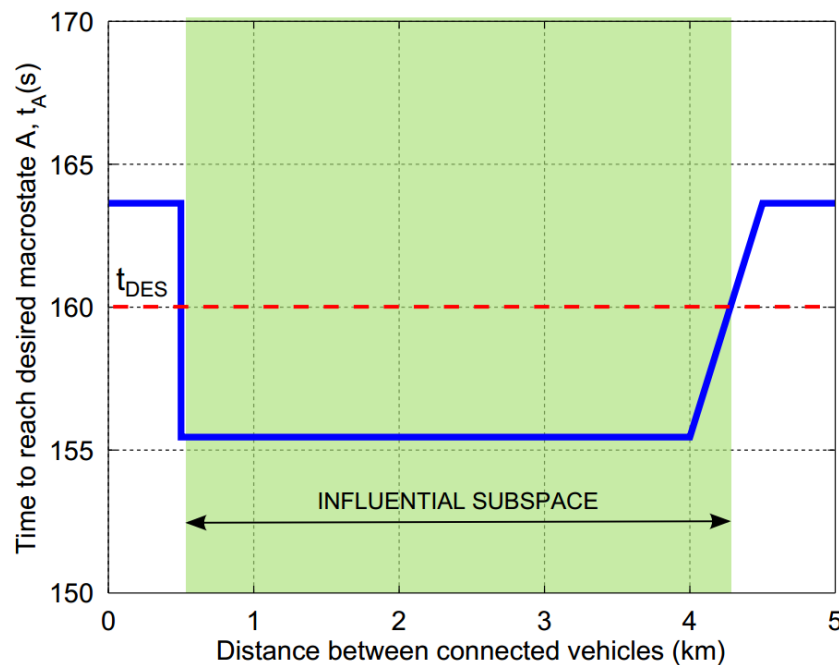


FIGURE 5 Influential subspace of the second connected vehicle, given the desired macrostate A and pre-specified time t_{DES} .

REFERENCES

1. Research and Innovative Technology Administration, U.S. Department of Transportation. *Connected Vehicle Research*. www.its.dot.gov/connected_vehicle/connected_vehicle.htm. Accessed March 24, 2014.
2. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. In *Proceedings of the Royal Society A*, Vol. 229, No. 1178, 1955, pp. 317–345.
3. Richards, P. Shock Waves on the Highway. *Operations Research*, Vol. 4, No. 1, 1956.
4. Jerath, K., and S. N. Brennan. Analytical Prediction of Self-organized Traffic Jams as a Function of Increasing ACC Penetration. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 4, 2012, pp. 1782–1291.
5. Nishi, R., A. Tomoeda, K. Shimura, and K. Nishinari. Theory of Jam-Absorption Driving. *Transportation Research Part B*, Vol. 50, 2013, pp. 116–129.
6. Shladover, S., B. van Arem, and B. D. Netten. *Effects of Cooperative Adaptive Cruise Control on Traffic Flow*. California PATH Program, University of California, Berkeley, 2009.
7. Monteil, J., R. Billot, J. Sau, F. Armetta, S. Hassas, S., and N.-E. El Faouzi. Cooperative Highway Traffic: Multiagent Modeling and Robustness Assessment of Local Perturbations. In *Transportation Research Record: Journal of the Transportation Research Record*, No. 2391, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 1–10.

FREEWAY TRAFFIC ANALYSIS

The Heterogeneity of Capacity Distributions Among Different Freeway Lanes

KUN XIE
KAAN OZBAY
HONG YANG
New York University

The stochastic nature of highway capacity has drawn increasing attention in recent studies, where the capacity distributions are usually estimated for whole freeway sections using aggregated data from multiple lanes. However, those cross section–based capacity models cannot be used to assess the risk of semicongested states observed, in which traffic breaks down on certain lanes while still flowing uninterrupted on others. This study attempts to investigate the freeway capacity heterogeneity among individual lanes using robust statistical estimation methods. Four diverge sections of different interstate highways in California were selected for the study. Optimal threshold speed was identified for each lane by maximizing the average reduction of traffic efficiency, which was defined as the product of mean speed and volume. Capacity observations were obtained for individual lanes, and censoring was indicated, both based on the optimal threshold speeds. Log-rank and Wilcoxon tests were conducted, and the results confirmed the heterogeneity of capacity distributions among lanes of the same freeway section and showed the necessity of estimating capacity distributions for individual lanes separately. A Bayesian hierarchical Weibull model based on censored capacity data was used to estimate these lane-specific capacity distributions. The model parameters are allowed to vary across freeway sections to account for unobserved heterogeneity and accordingly to improve the accuracy of estimations. In addition, censored data issues are adequately addressed in the proposed model. It is found that breakdown probability would be overestimated if censoring is ignored. The proposed model can provide useful insights when diagnosing bottlenecks with semicongested cases, which may be neglected by the cross section–based models.

INTRODUCTION

Traditional highway capacity is defined as the maximum traffic flow rate that traverses a section under prevailing roadway, traffic, and control conditions (1). However, the use of deterministic traffic capacity is an incomplete representation of real-life conditions. The concept of stochastic highway capacity has been recently discussed in several previous studies (2–4), where the capacity is defined as the traffic volume below which traffic still flows and above which the flow breaks down. The capacity in this sense can be regarded as a variable, since the flow rate that causes traffic breakdown is related to a variety of factors such as traffic composition, driving behavior, and environmental characteristics (5). The study by Persaud et al. (6) on the probabilistic breakdown phenomenon in freeway traffic convincingly showed that traffic capacity is stochastic in nature.

Regarding the stochastic feature of freeway capacity, it is important to use statistically robust methods to estimate its distribution based on the observed data. Brilon et al. (4) applied a nonparametric Kaplan-Meier estimator to obtain breakdown probability based on an analogy with the statistics of survival analysis. Censored observations on capacity were used in their studies. However, a complete capacity distribution function couldn't be obtained using the Kaplan-Meier due to the limitation of data points observed. In the following study by Brilon et al. (7), a parametric method was used to estimate the whole capacity distribution function, where the Weibull model appeared to be the best fit. Recently, Ozguven and Ozbay (8) proposed a nonparametric Bayesian survival analysis approach for the estimation of the freeway capacity. Their results indicated that Bayesian estimator could provide robust estimation outputs in the presence of insufficient and unreliable data.

Most of the previous studies (4, 7, 8) treated a multilane section as a single analysis unit, and aggregated data of all the lanes were used to identify breakdowns and estimate cross section-based capacity distributions. However, it should be noted that for multilane freeways, the traffic compositions and operational characteristics vary across different lanes. The vehicles with higher speed and better acceleration are more likely to travel on the median lanes. Moreover, the traffic flows of the shoulder lanes have a larger chance of being disturbed by merging and diverging vehicles on- and off-ramps. On the other hand, different drivers choose to travel on different lanes according to their destinations (9). In the study by Lawson et al. (10), higher occupancies were observed on the shoulder lanes, which were upstream of a congested off-ramp compared with the median lanes. Duret et al. (11) affirmed that traffic demand was not assigned evenly among lanes on a freeway section. Therefore, breakdown occurrence can be different even for the lanes of the same freeway section. An empirical study by Dehman and Drakopoulos (12) showed the significant difference of breakdown phenomena among lanes. Munoz and Daganzo (9) also noted a semi-congested state in which some lanes are congested, and other lanes of the same freeway section are not. The application of capacity models at the cross-section level will thus ignore the high breakdown probability of certain lanes and cannot accurately assess the risk of semicongestions. Ma et al. (13) proposed a lane-based method to identify breakdown for diverge sections, but capacity distributions for specific lanes were not estimated. In addition, censored data and unobserved heterogeneity across different freeway sections were not considered in their study.

This study attempts to investigate the capacity heterogeneity among individual lanes of diverge sections using robust statistical methods. Semicongestion is a common phenomenon in diverge sections, where the traffic is likely to breakdown on the shoulder lanes connected to the off-ramps, even though the traffic on the other lanes is still flowing. This paper starts with data preparation, where capacity observations and censored data are collected for specific lanes based on optimal threshold speed identified. Hypothesis tests for the heterogeneity of capacity distributions among lanes are conducted to confirm the necessity of developing capacity models at the individual lane level. After that, a novel hierarchical Weibull model is proposed in the Bayesian framework to estimate lane-specific capacity distributions. The parameters of the model are allowed to vary across various sections to account for unobserved heterogeneity. In addition, censored data issues described in Brilon et al. (4) and Ozguven and Ozbay (8) are explicitly addressed in the proposed model.

DATA PREPARATION

Traffic data used in this study were obtained from the Performance Measurement System (PeMS) of the California Department of Transportation. Four diverge sections that have vehicle detector stations (VDS) right upstream of the off-ramps were selected. These sections have the same lane configuration as shown in Figure 1, but they are located in different interstate highways, and thus their traffic and environmental features are different. The basic information of the diverge sections selected is presented in Table 1. Traffic speed and volume data of the selected sections from January 1 to January 31 in 2014 were extracted for each individual lane.

Capacity Observations and Censoring

A traffic breakdown is usually followed by a significant amount of reduction in speed. A common method [first proposed in Brilon et al. (4) and later applied in Brilon et al. (7), Ozguven and Ozbay (8), and Ma et al. (13)] to identify traffic breakdown points based on mean speed was used in this study. The mean speed (mph) and volume (veh/h) were calculated for each lane at 5-min intervals. The traffic volume of interval i was taken as an observation of capacity if the mean speed starts to fall below threshold speed in the next interval $i + 1$.

Censoring refers to a form of imperfect but useful observation of a variable. Censored data are, however, valuable but require special consideration to avoid biased estimation results (14, 15). Ignoring censored data could lead to biased estimates of capacity distributions, especially for cases with inadequate breakdown points identified. The identification method of censored data adopted here is consistent with the previous studies such as Brilon et al. (4) and Ozguven and Ozbay (8). If mean speed is higher than the threshold speed in the current interval i and the following interval $i + 1$, the volume of the interval i was taken as a right-censored observation. A right-censored observation provides information that the actual capacity is ensured to be greater than the observed one. As illustrated in Figure 2, the volume of the fourth interval is taken as an observation of capacity, because the speed starts to be lower than the threshold speed in the following intervals. The volumes in the first three intervals are used as censored observations, since these observed volumes are not high enough to cause breakdowns. Traffic is in the congested state in the last three intervals. As suggested by Brilon et al. (4), the volumes of these intervals don't provide any information for the capacity distribution and thus are excluded from the model development.

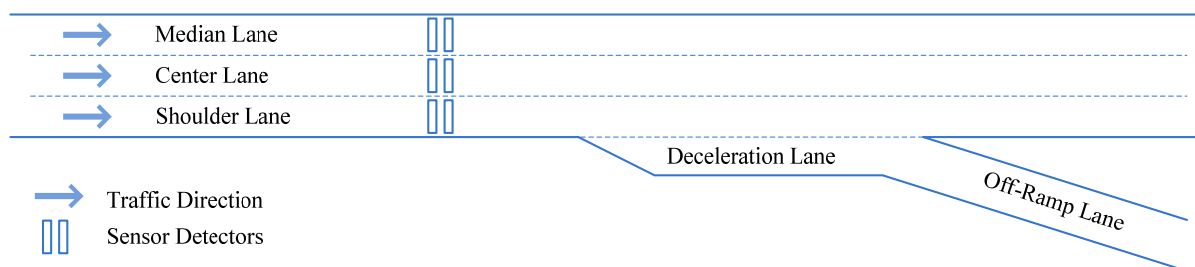


FIGURE 1 Lane configuration and sensor detector location of the four diverge sections selected.

TABLE 1 Overview of the Selected Diverge Sections

	VDS ID	Highway	Post Mile	Direction	County
Section 1	715918	I-5	122.8	Northbound	Los Angeles
Section 2	716448	I-105	12.1	Westbound	Los Angeles
Section 3	819634	I-15	91.89	Northbound	Riverside
Section 4	809070	I-210	61.14	Eastbound	San Bernardino

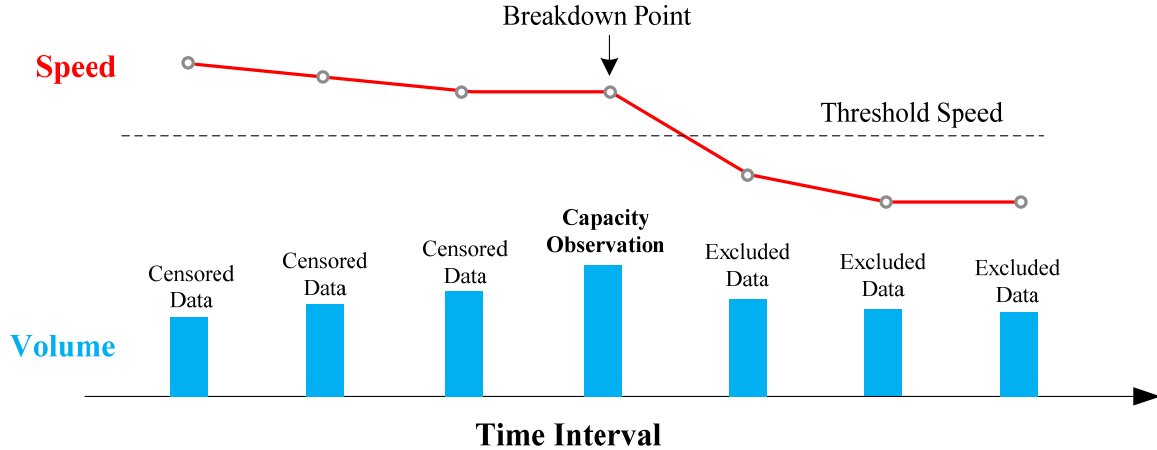


FIGURE 2 Capacity observation and censored data during a breakdown.

Optimal Threshold Speed

In previous research (4, 7, 8), a standard threshold speed of 70 km/h (about 43.75 mph) was used. However, it is undesirable to use the same threshold speed for different lanes, because the threshold speed is likely to be associated with geometric design, traffic composition, and environmental features. Ma et al. (13) introduced a criterion that can identify the optimal threshold speed by maximizing mean speed reduction and used it to determine threshold speeds for specific lanes. Noting that a traffic breakdown is always accompanied by a significant reduction not only in the traffic speed but also in the traffic volume, the authors propose to use the product of speed and volume to identify optimal threshold speed in this study.

Let v_i and q_i denote the speed and volume at interval i . A measure called traffic efficiency, E_i , can be calculated as the product of v_i and q_i , namely $E_i = v_i \times q_i$. The concept of traffic efficiency was first introduced by Brilon (16). For a certain threshold speed $V_{Threshold}$, if v_i is greater than $V_{Threshold}$ and the speed in the next interval v_{i+1} is lower than $V_{Threshold}$, a breakdown is identified that is accompanied by a reduction of efficiency $\Delta E_i = v_i \times q_i - v_{i+1} \times q_{i+1}$. For all the breakdown points identified using $V_{Threshold}$, the average of efficiency reduction $\overline{\Delta E_i}$ can be computed. After obtaining $\overline{\Delta E_i}$ using a variety of $V_{Threshold}$, the optimal threshold speed $V_{Threshold}^*$ can be identified as the one that leads to the maximum $\overline{\Delta E_i}$.

The aforementioned optimization process is used to compute the optimal threshold

speeds for specific lanes and cross sections. As presented in Table 2, the average optimal threshold speed for cross sections is 44.00 mph, which is quite close to the standard threshold speed of 43.75 mph. The data of the median lane in Section 4 are used to demonstrate the breakdown points identified by the standard threshold speed (43.75 mph) and the optimal threshold speed (52.00 mph), as shown in Figure 3. The results show that it is more appropriate to use the optimal threshold speed to determine breakdown occurrence since the average of capacity observations identified by the optimal threshold speed (1,999 veh/h/lane) is higher than that obtained by the standard threshold speed (1,926 veh/h/lane). The optimal threshold speeds listed in Table 2 were used to extract the capacity observations and indicate censoring for individual lanes and cross sections.

TABLE 2 Optimal Threshold Speeds for Specific Lanes and Cross Sections (mile/h)

	Median Lane	Center Lane	Shoulder Lane	Cross Section
Section 1	41	35	33	39
Section 2	46	40	37	41
Section 3	49	43	40	48
Section 4	52	47	40	48
Average	47.00	41.25	38.00	44.00

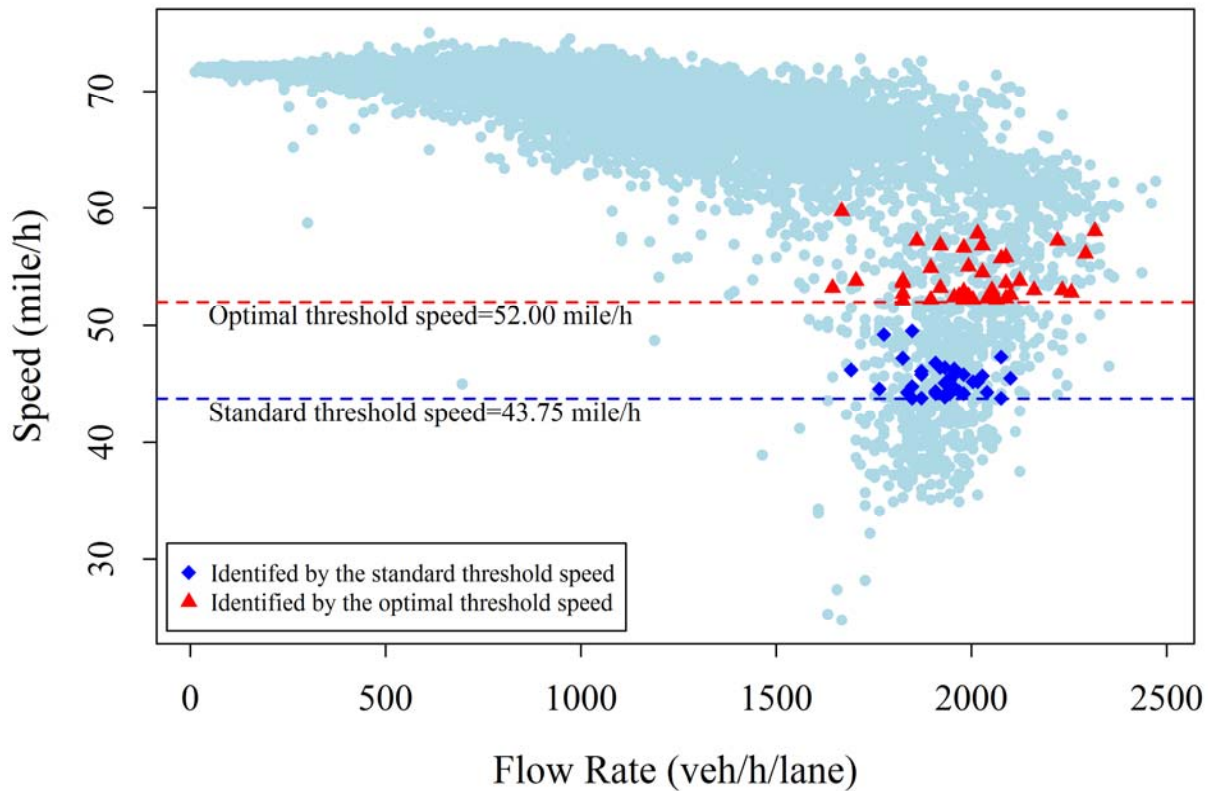


FIGURE 3 Breakdown points identified by standard and optimal threshold speeds (using the median lane of Section 4 for demonstration).

HYPOTHESIS TESTS FOR CHECKING CAPACITY HETEROGENEITY

Hypothesis tests for heterogeneity of capacity distributions across lanes were performed using censored capacity data. Let $h_k(q)$ denote the breakdown hazard rate at volume q for the k th lane ($k = 1, 2, \dots, K$). To compare the capacity distributions among different lanes, the null and alternative hypotheses are set as follows:

$$H_0: h_1(q) = h_2(q) = \dots = h_K(q) \text{ for all } q.$$

$$H_A: h_{k_1}(q) \neq h_{k_2}(q) \text{ for at least one pair of } k_1 \text{ and } k_2, \text{ where } 1 \leq k_1 \neq k_2 \leq K.$$

The volume observations from different lanes were pooled and sorted in ascending order, as follows $q_1 \leq q_2 \leq \dots \leq q_D$. At volume q_i , the authors observed d_{ik} breakdowns from the k th lane out of Y_{ik} individuals at risk of breakdown, where $i = 1, 2, \dots, D$. The Nelson-Aalen method (17) provides a nonparametric estimator of the hazard rate using censored data. Based on the Nelson-Aalen method, the quantity d_{ik}/Y_{ik} gives an estimate of breakdown hazard rate at volume q_i for the k th lane. It should be noted that the censored observations are supposed to be counted in obtaining Y_{ik} . Similarly, under the null hypothesis, the pooled breakdown hazard rate at volume q_i for the cross section is d_i/Y_i , where $d_i = \sum_{k=1}^K d_{ik}$ and $Y_i = \sum_{k=1}^K Y_{ik}$. The test of H_0 is based on the weighted difference of estimated hazard rates (18):

$$Z_k = \sum_{i=1}^D W_k(q_i) \left(\frac{d_{ik}}{Y_{ik}} - \frac{d_i}{Y_i} \right) \quad (1)$$

Where $W_k(q_i)$ is the weight function for the k th lane at volume q_i . In practice, since the commonly used weight functions have the form $W_k(q_i) = Y_{ik}W(q_i)$, then the following exists:

$$Z_k = \sum_{i=1}^D W(q_i) (d_{ik} - Y_{ik} \frac{d_i}{Y_i}) \quad (2)$$

Greater Z_k indicates that the breakdown hazard rate of the k th lane tends to be more different from the expected one under the null hypothesis.

The variance of Z_k and the covariance of Z_{k_1} and Z_{k_2} are determined by (18)

$$\sigma_k = \sum_{i=1}^D W(q_i)^2 \frac{Y_{ik}}{Y_i} \left(1 - \frac{Y_{ik}}{Y_i} \right) \left(\frac{Y_i - d_i}{Y_i - 1} \right) d_i \quad (3)$$

and

$$\sigma_{k_1 k_2} = - \sum_{i=1}^D W(q_i)^2 \frac{Y_{ik_1}}{Y_i} \frac{Y_{ik_2}}{Y_i} \left(\frac{Y_i - d_i}{Y_i - 1} \right) d_i \quad (4)$$

Let Σ represents the estimated $(K - 1) \times (K - 1)$ variance–covariance matrix of Z_k , where $k = 1, 2, \dots, K - 1$. The overall test statistic for the hypothesis is then defined as follows (18):

$$\chi^2 = (Z_1, Z_2, \dots, Z_{K-1})\Sigma^{-1}(Z_1, Z_2, \dots, Z_{K-1})' \tag{5}$$

This overall test statistic is treated as a chi-square distribution with degree of freedom equal to $K - 1$. To calculate the χ^2 , a variety of weight functions are proposed in the literature. If the $W(q_i) = 1$ is selected, it leads to the log-rank test (19), where greater weights are assigned to larger observations. If the weight function is set to be $W(q_i) = Y_i$, it results in the generalized Wilcoxon test (20), where smaller observations are heavily weighted.

Both the log-rank and Wilcoxon tests were conducted to assess the across-lane heterogeneity of capacity distributions for each section separately, based on censored capacity data. As presented in Table 3, p -values from both the log-rank and the Wilcoxon tests are less than 0.0001, so the null hypothesis H_0 should be rejected and the alternative hypothesis H_A are accepted. The results indicate significant heterogeneity of capacity distributions among different lanes. To improve the accuracy of assessing breakdown risk, it is critical to develop lane-specific capacity models, which are specified in the next section.

METHODOLOGY OF CAPACITY ESTIMATION

In this section, a full Bayesian approach is proposed to estimate the lane-specific capacity distributions. A Bayesian approach provides the ability to deal with insufficient data issues, to flexibly select parameter distributions, and to accommodate complicated model structures (14, 21). In Bayesian models, prior distributions are combined with a likelihood function obtained from the observed data to estimate posterior distributions. The Bayesian estimation procedure, prior distributions, and assessment of Bayesian models are also introduced in this section.

Bayesian Hierarchical Weibull Model

Let $q = (q^1, q_2, \dots, q_n)$ denote the traffic volume data. According to the previous research (7, 13), q can be assumed to follow a Weibull distribution $w(\alpha, \lambda)$. The probability density function and cumulative density function of q are given by

TABLE 3 Results of Log-Rank and Wilcoxon Tests

	Test	Chi-Square	Degree of Freedom	P-Value	Accepted Hypothesis
Section 1	Log-rank	5180	2	<0.0001	Reject H_0 , accept H_A
	Wilcoxon	2616	2	<0.0001	Reject H_0 , accept H_A
Section 2	Log-rank	268	2	<0.0001	Reject H_0 , accept H_A
	Wilcoxon	71	2	<0.0001	Reject H_0 , accept H_A
Section 3	Log-rank	3743	2	<0.0001	Reject H_0 , accept H_A
	Wilcoxon	1309	2	<0.0001	Reject H_0 , accept H_A
Section 4	Log-rank	45	2	<0.0001	Reject H_0 , accept H_A
	Wilcoxon	52	2	<0.0001	Reject H_0 , accept H_A

$$f(q | \alpha, \lambda) = \alpha q^{\alpha-1} \exp(\lambda - \exp(\lambda)q^\alpha) \quad (6)$$

and

$$F(q | \alpha, \lambda) = \int f(q | \alpha, \lambda) dq = 1 - \exp(-\exp(\lambda)q^\alpha) \quad (7)$$

The survival function $S(q | \alpha, \lambda)$ gives the probability that traffic keeps flowing until it reaches volume q . The $S(q | \alpha, \lambda)$ can be derived from $F(q | \alpha, \lambda)$ as follows:

$$S(q | \alpha, \lambda) = 1 - F(q | \alpha, \lambda) = \exp(-\exp(\lambda)q^\alpha) \quad (8)$$

To quantify the instantaneous risk that traffic breaks down at volume q , a hazard function $H(q | \alpha, \lambda)$ is introduced. $H(q | \alpha, \lambda)$ is defined as the conditional probability of breakdown:

$$H(q | \alpha, \lambda) = \frac{f(q | \alpha, \lambda)}{S(q | \alpha, \lambda)} = \alpha q^{\alpha-1} \exp(\lambda) \quad (9)$$

According to Equation 9, if $\alpha > 1$, then the breakdown hazard increases as volume q increases; if $\alpha < 1$, then the breakdown hazard is negatively associated with volume q ; if $\alpha = 1$, then the breakdown hazard is not related to volume q .

Let $v = (v_1, v_2, \dots, v_n)'$ denote the censoring indicators, where $v_i = 1$ indicates q_i is a breakdown volume, and $v_i = 0$ indicates q_i is a censored observation. A censored observation expresses the information that the actual capacity is ensured to be greater than the observed volume. Including the censored observations, the likelihood function of (α, λ) can be written as

$$\begin{aligned} L(\alpha, \lambda | q, v) &= \prod_{i=1}^n f(q_i | \alpha, \lambda)^{v_i} S(q_i | \alpha, \lambda)^{1-v_i} \\ &= \alpha^d \exp\{d\lambda + \sum_{i=1}^n (v_i(\alpha-1) \log(q_i) - \exp(\lambda)q_i^\alpha)\} \end{aligned} \quad (10)$$

where $d = \sum_1^n v_i$. For the Bayesian Weibull model, it is usually assumed that the parameter α follows a gamma prior distribution $\Gamma(r_0, m_0)$ and λ follows a normal distribution $N(\mu_0, \sigma_0^2)$ (14). The joint posterior distribution of (α, λ) is given by (14):

$$\begin{aligned} \pi(\alpha, \lambda | q, v) &\propto L(\alpha, \lambda | q, v) \pi(\alpha | r_0, m_0) \pi(\lambda | \mu_0, \sigma_0^2) \\ &\propto \alpha^{r_0+d-1} \exp\{d\lambda + \sum_{i=1}^n (v_i(\alpha-1) \log(q_i) - \exp(\lambda)q_i^\alpha) - m_0\alpha - \frac{1}{2\sigma_0^2}(\lambda - \mu_0)^2\} \end{aligned} \quad (11)$$

To estimate lane-specific capacity distributions, different sets of Weibull parameters should be assigned to different lanes. In addition, the data structure used in this study can be viewed as a two-level hierarchy with Level 1 being the lane level, and Level 2 being the freeway section level. Accordingly, a hierarchical model structure is proposed where parameters are allowed to vary to account for the unobserved heterogeneity across sections (22). Weibull parameters $(\alpha_{kj}, \lambda_{kj})$ for the k th lane at the j th section can be specified as follows:

$$\alpha_{kj} \sim \Gamma(r_k, m_k) \quad (12)$$

and

$$\lambda_{kj} \sim N(\mu_k, \sigma_k^2) \quad (13)$$

In Equation 6, α_{kj} and λ_{kj} are taken as random parameters that can vary across the sections. Parameters r_k , m_k , μ_k , and σ_k^2 are specific for each lane. For the k th lane, the mean and variance of α_{kj} are r_k/m_k and r_k / m_k^2 ; the mean and variance of λ_{kj} are μ_k and σ_k^2 .

Bayesian Estimation Procedure and Priors

Bayesian models are usually estimated via a Markov Chain Monte Carlo (MCMC) algorithm (23). The primary technique of MCMC is Gibbs sampling (24), each iteration of which draws a new value for each unobserved stochastic node from its full conditional distribution given the current values of all the other quantities in the model (25). The WinBUGS statistical software package was used to provide a computing approach for the calibration of Bayesian models using Gibbs sampling (26).

In the absence of credible prior information for model parameters, uninformative priors were used to express vague and general information about parameters. The priors of r_k and m_k were assumed to be the lognormal distribution $(0, 10^5)$; μ_k was assumed to follow the normal distribution $(0, 10^5)$, and σ_k^2 was assumed with the Inverse-Gamma distribution $(10^{-3}, 10^{-3})$. Considering convergence and time of updating, two MCMC chains of 30,000 iterations were run, and the first 10,000 samples were discarded as burn-in. The Brooks-Gelman-Rubin (BGR) diagnostic proposed by Brooks and Gelman (27) was used to assess the convergence of multiple chains. Convergence was assumed to occur when the BGR statistic is less than 1.2.

Deviance Information Criterion

The deviance information criterion (DIC) offers a Bayesian measure of model fitting and complexity (26). Specifically, DIC is calculated as follows:

$$DIC = \overline{D(\theta)} + p_D \quad (14)$$

where $\overline{D(\theta)}$ denotes the posterior mean of Bayesian deviance of parameter θ and can be used to indicate how well the model fits the data. p_D defines the effective number of parameters and is taken as a measure of model complexity. Models with smaller DIC are preferred.

RESULTS AND DISCUSSION

The proposed Bayesian hierarchical Weibull model structure was used to estimate lane-specific capacity distributions. To conduct comparisons, the authors also developed standard Weibull models where parameters α and λ were fixed for different freeway sections. In addition, to study the impacts of censoring on estimates, two datasets were prepared for modeling: uncensored dataset, where the censored observations on capacity were excluded; and censored dataset, which included censored observations. In view of the long convergence time to estimate Bayesian models with large amounts of data, the authors sampled the censored observations of the second dataset for the estimation process. The hierarchical Weibull and standard Weibull models were developed based on these two datasets, respectively. Bayesian posterior estimates of parameters and DICs of those models are reported in. $sd(\lambda)$ and $sd(\alpha)$ represent the standard deviations of parameter distributions across sections in the hierarchical models. It should be noted that $sd(\lambda)$ and $sd(\alpha)$ are unavailable for standard Weibull models in which parameters are fixed for different sections. The 95% Bayesian Credible Interval (95% BCI) was used to examine the significance of estimates. Estimates can be regarded as significant at the 95% level if the BCIs do not cover 0 and vice versa (28). Parameter estimates are all found to be significant, as their 95% BICs do not cover 0.

Benefits of Hierarchical Structure and Explicitly Addressing Censoring

As shown in Table 4, for the uncensored dataset, $\overline{D(\theta)}$ of the hierarchical Weibull model (9,610) is 522 less than that of the standard Weibull model (10,132). Similarly, based on the censored dataset, the hierarchical Weibull model has a lower $\overline{D(\theta)}$ by 423 in comparison with the standard Weibull model. These findings indicate that the hierarchical Weibull models outperform the standard Weibull models in term of the goodness of fit. In another aspect, higher p_D values are observed in the hierarchical Weibull models. This reflects the increasing complexity as a result of including random parameters. Overall, regarding the lower DICs, the hierarchical Weibull models show substantial improvement, although they are penalized by higher p_D values. Note that it is meaningless to compare models using DIC, unless they are developed based on the same dataset. So it is erroneous to conclude that the hierarchical model developed based on the uncensored dataset is superior to the one based on censored dataset, although the former has a smaller DIC. The hierarchical models allow parameters α and λ to vary across sections and thus have the flexibility to account for the unobserved heterogeneity. All the estimates of $sd(\lambda)$ and $sd(\alpha)$ are found to be significantly different from 0, and this provides strong evidence for the presence of heterogeneity among sections. By capturing the unobserved heterogeneity, the effects of parameters can be adjusted. As a result the estimates of α and λ of hierarchical Weibull models differ significantly from those of the standard Weibull models. On the other hand, for each section, the same set of parameters is used to duplicate the capacity distributions in the hierarchical Weibull models, and in this sense, the within-section correlation of capacity observations can be addressed simultaneously.

Using the estimated parameters and Weibull cumulative density function shown in Equation 7, the 25th, 50th, and 75th percentiles of capacity distributions were calculated as presented in Table 5. Based on the hierarchical Weibull models (uncensored or censored dataset), the 25th, 50th, and 75th percentiles of the median lanes are greater than those of the center lanes, respectively. This indicates that greater flow rate is needed for the median lanes

TABLE 4 Posterior Summary of Bayesian Models

	Standard Weibull		Hierarchical Weibull	
	Uncensored Dataset	Censored Dataset	Uncensored Dataset	Censored Dataset
	Mean (95% BCI)	Mean (95% BCI)	Mean (95% BCI)	Mean (95% BCI)
Median Lane				
λ	-60.05 (-66.08, -55.64)	-65.89 (-71.46, -59.76)	-76.48 (-83.25, -69.13)	-70.98 (-76.11, -63.67)
$sd(\lambda)$	—	—	0.38 (0.02, 1.59)	0.56 (0.02, 2.65)
α	7.90 (7.33, 8.69)	8.64 (7.83, 9.36)	10.12 (9.09, 11.02)	9.33 (8.28, 10.29)
$sd(\alpha)$	—	—	0.13 (0.07, 0.27)	0.12 (0.04, 0.38)
Center Lane				
λ	-60.09 (-64.92, -55.05)	-64.7 (-70.47, -58.37)	-73.91 (-81.26, -67.48)	-72.73 (-75.79, -67.29)
$sd(\lambda)$	—	—	0.59 (0.02, 3.2)	0.40 (0.02, 1.64)
α	8.04 (7.37, 8.68)	8.61 (7.76, 9.37)	10.00 (8.74, 11.06)	9.75 (8.94, 10.24)
$sd(\alpha)$	—	—	0.17 (0.08, 0.45)	0.11 (0.05, 0.21)
Shoulder Lane				
λ	-44.68 (-49.15, -40.57)	-50.45 (-55.33, -45.4)	-85.6 (-93.42, -76.22)	-75.16 (-80.96, -67.09)
$sd(\lambda)$	—	—	1.29 (0.02, 5.44)	1.36 (0.02, 5.72)
α	5.96 (5.41, 6.55)	6.69 (6.02, 7.33)	11.65 (10.3, 12.7)	10.15 (9.13, 10.96)
$sd(\alpha)$	—	—	0.32 (0.11, 0.62)	0.27 (0.08, 0.54)
Assessment				
$\bar{D}(\theta)$	10,132	10,543	9,610	10,120
p_D	6	6	16	17
DIC	10,138	10,549	9,626	10,137

than for the center lanes to result in the same breakdown probability. Similarly, as anticipated, the center lanes have higher percentiles compared with the shoulder lanes. However, some inconsistent results (marked with an asterisk) are observed when capacity distributions are estimated using the standard Weibull models. The 50th and 75th percentiles of the shoulder lanes are found to be even greater than those of the center lanes, respectively. This result further confirms the advantage of the hierarchical Weibull models over the standard Weibull models in adjusting the parameter estimates. In addition, as shown in Table 5, the models developed based on censored dataset yield greater percentiles than those excluding censored observations. This finding implies that breakdown probability would be overestimated if censoring is ignored. To get more reliable statistic inferences, the censored observations have to be treated properly.

Lane-Specific Capacity Distributions

In view of its superiority, the hierarchical Weibull model developed based on censored dataset was selected to estimate the capacity distributions of specific lanes. As a comparison, the hierarchical Weibull model was also used to estimate the capacity distribution for cross sections ($\alpha = 10.80, \lambda = -80.84$) based on aggregated volume and speed data of multiple lanes.

TABLE 5 The 25th, 50th, and 75th Percentiles of the Estimated Capacity Distributions (veh/h/lane)

Standard Weibull	Uncensored Dataset			Censored Dataset		
	Median Lane	Center Lane	Shoulder Lane	Median Lane	Center Lane	Shoulder Lane
25th percentile	1,702	1,506	1,469	1,783	1,591	1,572
50th percentile	1,902	1,680	1,702 ^a	1,974	1,762	1,793 ^a
75th percentile	2,077	1,832	1,913 ^a	2,139	1,910	1,989 ^a
Hierarchical Weibull	Uncensored Dataset			Censored Dataset		
	Median Lane	Center Lane	Shoulder Lane	Median Lane	Center Lane	Shoulder Lane
25th percentile	1,692	1,434	1,395	1,759	1,526	1,454
50th percentile	1,846	1,566	1,504	1,932	1,670	1,585
75th percentile	1,977	1,678	1,596	2,082	1,794	1,697

^aInconsistent results.

Substituting the estimates of α and λ into the Weibull cumulative density function given in Equation 7 and the Weibull hazard function shown in Equation 9, breakdown probability and hazard curves for specific lanes and cross sections are plotted. As shown in Figure 4, the shoulder lanes have the highest expected breakdown probability and hazard, while the median lanes have the lowest ones, given the same traffic flow rate. According to Figure 4a, when the traffic demand is approaching 2,000 veh/h/lane, the shoulder and center lanes have extremely high probability to break down, whereas for the median lanes the breakdown probability is approximately 0.6. In Figure 4b, all the hazard curves are increasing monotonically, which means that traffic is more likely to break down as the volume increases.

The breakdown probability and hazard curves of the cross sections are close to those of the center lanes but are significantly different from those of the median and shoulder lanes. It is found that using the cross section–based capacity models for individual lanes would underestimate the breakdown probability and hazard for the shoulder lanes and conversely overestimate them for the median lanes. The lane-based capacity models can be used to estimate the breakdown probability for specific lanes and thus are helpful in improving the accuracy of assessing the congestion risk, especially for the semicongested cases; however, the high breakdown probability in certain lanes would be ignored by section-based capacity models.

Using the proposed Bayesian hierarchical model, Weibull parameters for each diverge section can also be estimated. As shown in Figure 5, the breakdown probability curves differ from one section to another, and this result is consistent with the significance of $sd(\lambda)$ and $sd(\alpha)$. Moreover, the shoulder lanes have higher $sd(\lambda)$ and $sd(\alpha)$ than those of the median and center lanes according to. This indicates that the capacity distributions of the shoulder lanes vary greatly among sections. It is shown in Figure 5 that the lane-specific capacity distributions of Section 3 have greater variability than those of others, whereas the capacity distributions of

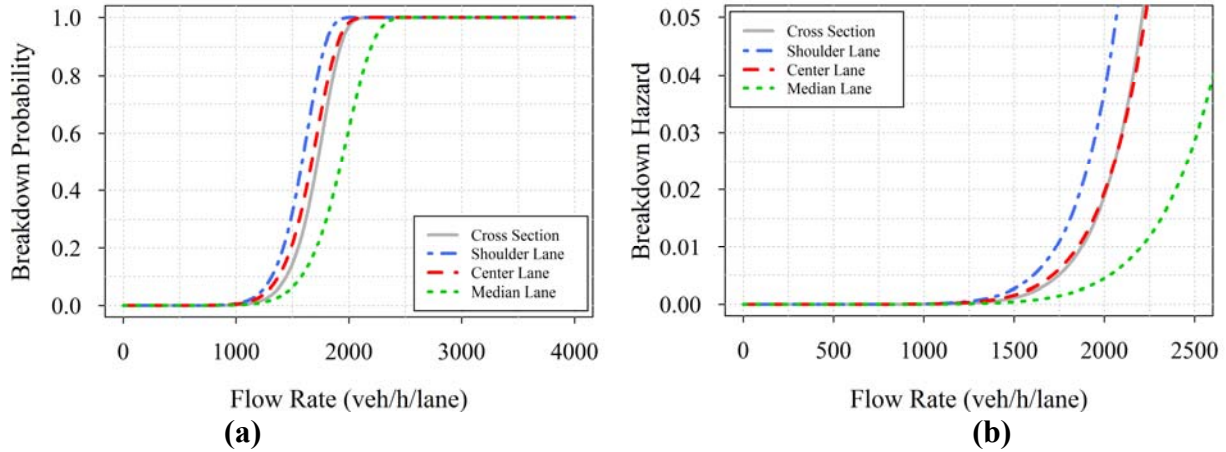


FIGURE 4 Breakdown probability and hazard curves for specific lanes and cross sections based on pooled data: (a) breakdown probability and (b) breakdown hazard.

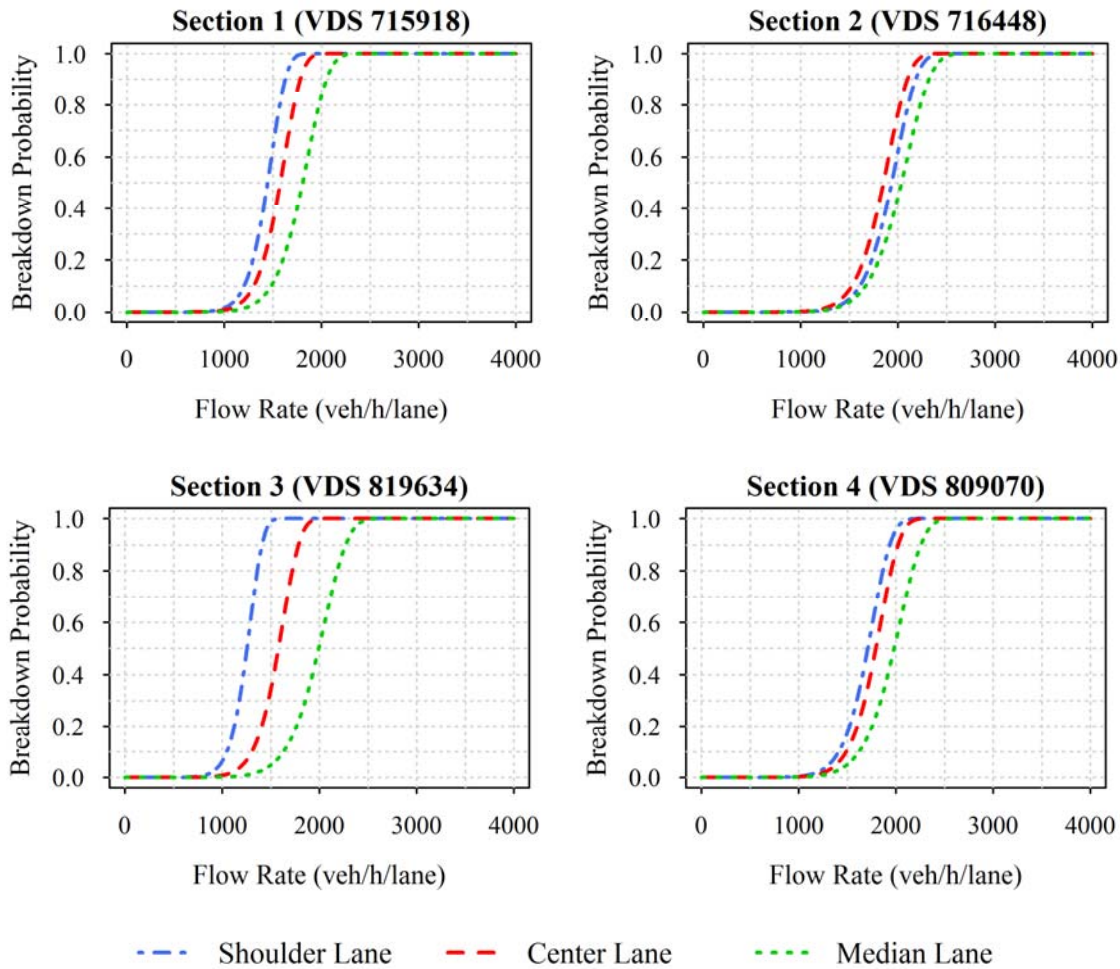


FIGURE 5 Lane-specific breakdown probability curves for each section.

different lanes in Sections 2 and 4 tend to be closer to each other. A counterintuitive result is that the breakdown probability of the shoulder lane of Section 2 is slightly lower than that of the center lane given the same traffic flow rate. A possible reason is that Section 2 provides proper alignment with the off-ramp, and thus the interaction impacts on the traffic of the shoulder lane may be smaller. It may also be attributed to the different traffic compositions and driving behaviors of specific lanes.

CONCLUSIONS

This study contributes to the literature of stochastic freeway capacity by estimating lane-specific capacity distributions using a novel Bayesian hierarchical Weibull model. Four diverge sections of different interstate highways in California were studied. The speed and volume data of those four sections were retrieved from the PeMS. Noting the difference of traffic features among lanes, a method to obtain the optimal threshold speed by maximizing the average reduction of efficiency was proposed and used to identify traffic capacity observations and indicate censoring. Higher breakdown volumes could be observed by using the optimal threshold speed compared with the standard threshold speed proposed in the literature. Prior to model development, the censored capacity data were examined by the log-rank and Wilcoxon tests. The test results provided strong evidences for the heterogeneity of capacity distributions among individual lanes of the same section and showed the necessity of modeling capacity distributions separately at the lane level. By comparing with other models, the hierarchical Weibull model based on the censored dataset had the best performance and thus was used to investigate the capacity distributions of specific lanes. It is found that breakdown hazard is positively associated with the prevailing traffic volume and that lane-specific capacity distributions differ from one section to another. The merits of the proposed model are listed as follows:

- Different sets of Weibull parameters are assumed for the median, center, and shoulder lanes. Such model specification can support lane-based capacity analysis by estimating the breakdown probability for each individual lane.
- The parameters of the hierarchical Weibull model are allowed to vary across freeway sections. By accounting for the across-section unobserved heterogeneity and the within-section correlation simultaneously, the parameter estimates are properly adjusted. In light of its lower DIC value, the proposed model shows substantial improvement over the standard Weibull model.
- By integrating the survival analysis theory in the proposed model, the censored data are appropriately treated. It is found that breakdown probability would be overestimated if censoring is ignored.
- A full Bayesian approach is adopted. It improves the proposed model in terms of addressing insufficient data issues, providing flexibility in selecting parameter distributions, and accommodating complicated model structures.

The proposed lane-specific capacity model can improve the accuracy of breakdown probability estimation and can provide useful insights when diagnosing bottlenecks with semicongested cases (i.e., congestions occurring only on certain lanes). Traffic operational performance can be improved by navigating vehicles to choose uncongested lanes to reduce the breakdown occurrence according to the outputs of the lanes-based capacity models. The findings

of this paper need further validation using larger data sets from freeway sections of geographic diversity. Additional work to compare Weibull distributions with other types of distributions in the Bayesian hierarchical framework is also suggested. The study on temporal correlation of capacity distributions is also of future interest.

ACKNOWLEDGMENTS

This study was partially supported by the Center for Urban Science and Progress (CUSP) and the Urban Mobility and Intelligent Transportation Systems (UrbanMITS) Laboratory at New York University. The authors thank the California Department of Transportation for providing the PeMS data.

REFERENCES

1. *Highway Capacity Manual*. TRB, National Research Council, Washington, D.C., 2000.
2. Lorenz, M. R., and L. Elefteriadou. A Probabilistic Approach to Defining Freeway Capacity and Breakdown. In *Transportation Research Circular E-C018: Fourth International Symposium on Highway Capacity*. TRB, National Research Council, Washington, D.C., 2000, pp 84–95.
3. Lorenz, M. R., and L. Elefteriadou. Defining Freeway Capacity as Function of Breakdown Probability. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1776*, TRB, National Research Council, Washington, D.C., 2001, pp. 43–51.
4. Brilon, W., J. Geistefeldt, and M. Regler. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. *Proc., 16th International Symposium on Transportation and Traffic Theory*, College Park, Maryland, 2005, pp. 125–144.
5. Elefteriadou, L., R. P. Roess, and W. R. McShane. Probabilistic Nature of Breakdown at Freeway Merge Junctions. In *Transportation Research Record 1484*, TRB, National Research Council, Washington, D.C., 1995, pp. 80–89.
6. Persaud, B., S. Yagar, and R. Brownlee, 1998. Exploration of the Breakdown Phenomenon in Freeway Traffic. In *Transportation Research Record 1634*, TRB, National Research Council, Washington, D.C., pp. 64–69.
7. Brilon, W., J. Geistefeldt, and H. Zurlinden. Implementing the Concept of Reliability for Highway Capacity Analysis. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2027*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 1–8.
8. Ozguven, E. E., and K. Ozbay. Nonparametric Bayesian Estimation of Freeway Capacity Distribution from Censored Observations. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2061*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 20–29.
9. Munoz, J. C., and C. F. Daganzo. The Bottleneck Mechanism of a Freeway Diverge. *Transportation Research Part A*, 36(6), 2002, pp. 483–505.
10. Lawson, T. W., W.-H. Lin, and M. J. Cassidy. *Validation of the Incremental Transfer Model*. California PATH Program, Institute of Transportation Studies, University of California at Berkeley, 1999.
11. Duret, A., S. Ahn, and C. Buisson. Lane Flow Distribution on a Three-Lane Freeway: General Features and the Effects of Traffic Controls. *Transportation Research Part C*, 24, 2012, pp. 157–167.
12. Dehman, A., and A. Drakopoulos. New Insights into Urban Freeway Breakdown Phenomenon. Presented at Mid-Continent Transportation Research Forum 2008, Madison, Wisconsin, 2008.

13. Ma, D., H. Nakamura, and M. Asano. Lane-Based Breakdown Identification at Diverge Sections for Breakdown Probability Modeling. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2395, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 83–92.
14. Ibrahim, J. G., M.-H. Chen, and D. Sinha. *Bayesian Survival Analysis*. Springer, New York, 2004.
15. Xie, K., K. Ozbay, and H. Yang. Before-After Safety Evaluations Using Parametric Bayesian Survival Analysis. Transportation Research Board 93rd Annual Meeting Compendium of Papers, 2014.
16. Brilon, W. Traffic Flow Analysis Beyond Traditional Methods. *Proc., 4th International Symposium on Highway Capacity*, 2000, pp. 26–41.
17. Nelson, W. Theory and Applications of Hazard Plotting for Censored Failure Data. *Technometrics*, 14(4), 1972, pp. 945–966.
18. Klein, J., and M. Moeschberger. *Survival Analysis: Statistical Methods for Censored and Truncated Data*. Springer-Verlag, New York, NY, 2003.
19. Mantel, N. Evaluation of Survival Data and Two New Rank Order Statistics Arising in Its Consideration. *Cancer Chemotherapy Reports*, Part 1 50(3), 1966, pp. 163–170.
20. Gehan, E. A. A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples. *Biometrika*, 52(1-2), 1965, pp. 203–223.
21. Xie, K., X. Wang, K. Ozbay, and H. Yang. Crash Frequency Modeling for Signalized Intersections in a High-Density Urban Road Network. *Analytic Methods in Accident Research 2*, 2014, pp. 39–51.
22. Xie, K., X. Wang, H. Huang, and X. Chen. Corridor-level Signalized Intersection Safety Analysis in Shanghai, China, Using Bayesian Hierarchical Models. *Accident Analysis and Prevention* 50, 2013, pp. 25–33.
23. Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall, Boca Raton, Fla., 1998.
24. Geman, S., and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6), 1984, pp. 721–741.
25. Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. WinBUGS: A Bayesian Modelling Framework: Concepts, Structure, and Extensibility. *Statistics and Computing* 10(4), 2000, pp. 325–337.
26. Spiegelhalter, D. J., N. G. Best, B. R. Carlin, and A. van der Linde. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 64, 2002, pp. 583–616.
27. Brooks, S. P., and A. Gelman. General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, 7(4), 1998, pp. 434–455.
28. Gelman, A. *Bayesian Data Analysis* (2nd ed.). Chapman & Hall/CRC, Boca Raton, Fla., 2004.

Calibration

CALIBRATION

Validating the Cost-Effectiveness Model for California's Freeway Incident Management Program

MICHAEL MAUCH
ALEX SKABARDONIS

Institute of Transportation Studies

LISA DAVIES

California Department of Transportation

Freeway service patrol (FSP) is a widely used incident management measure designed to assist disabled vehicles along congested freeway segments and reduce nonrecurring congestion through quick response to accidents and other incidents on freeways. An FSP beat performance evaluation model has been developed and used to analyze the cost-effectiveness of providing FSP service on selected freeway corridors and to assess the overall cost-effectiveness of California's FSP program. The FSP beat evaluation model estimates traffic delay savings, fuel savings, and emissions reductions per assisted incident as a result of FSP using deterministic queuing techniques.

This paper presents 1. a method of using real-world traffic and incident data to validate the FSP performance evaluation (FSPE) model and 2. the model validation results. The paper also presents key findings about the reliability freeway performance measures, like vehicle miles traveled (VMT), vehicle hours traveled (VHT), and traffic delays, estimated using the California Department of Transportation (Caltrans) Performance Measurement System (PeMS) stationary loop data and INRIX Analytics probe vehicle data.

INTRODUCTION AND BACKGROUND

Caltrans FSP is an incident management measure designed to assist disabled vehicles along congested freeway segments and reduce nonrecurring congestion through quick detection, response, and removal of accidents and other incidents on freeways. In California, the program is jointly administered by Caltrans, the California Highway Patrol (CHP), and regional transportation planning agencies. Currently, FSP operates on 193 freeway sites ("beats") across the state with 364 tow trucks over 1,800 centerline miles. California, having a large-scale FSP program and performance-driven decision-making policies, developed an analysis tool to evaluate the performance of FSP service on selected freeway corridors (i.e., FSP beats).

The benefits of providing FSP service depend on the beats' geometric and traffic characteristics and the frequency and type of assisted incidents. Incidents that occur in-lane tend to be more congestion-causing than shoulder incidents. Likewise, incidents occurring on freeways with high traffic demand (relatively little excess capacity) tend to cause more congestion than incidents on freeways with lower volumes. Earlier studies performed by the University of California, Berkeley (UC Berkeley), validated the FSP beat evaluation model by analyzing the effectiveness of FSP on a section of the I-880 freeway in the San Francisco Bay

Area (1) and a section of the I-10 freeway in Los Angeles (2). Extensive data on incidents and traffic characteristics were collected before and after the FSP deployment using specially instrumented probe vehicles and data from loop detectors. The data were processed, verified, and integrated into databases. Then analytical procedures were developed to estimate incident-specific delays. The resulting FSP performance evaluation (FSPE) model estimated benefits based on delay and fuel savings, fuel consumption, and air pollution reduction, and it showed that FSP was a cost-effective measure at the specific test sites.

These previous FSP model validation efforts focused on a very limited set of test sites, and previous model validation methodologies were applicable only to those FSP beats with relatively closely spaced PeMS vehicle detector stations. Methods that could be applied to a broader range of FSP beats (including FSP beats serving less-congested corridors and/or where PeMS detection stations are sparsely spaced or not available) would be better suited for statewide FSP model validation and performance monitoring purposes.

To address these needs, a method that was not dependent on tightly spaced (and fully functional) PeMS detector stations was developed to validate FSP delay savings for freeway corridors. The FSP's performance measures are directly derived from its vehicular delay savings; so, any validation method would need to quantify vehicular delays and delay savings attributable to FSP.

The next section of the paper (Section 2) introduces the concepts for the FSPE model. Section 3 discusses the methods used to validate the FSPE model. Section 4 introduces the data sources used to develop the validation targets, discusses data quality, and highlights some observations about the data. The results of the FSPE model validation efforts and an interpretation of the results follow in Section 5. Section 6, the last section, concludes with lessons learned and possible future work.

FSPE MODEL OVERVIEW

The FSPE model employs deterministic queuing techniques to estimate incident induced traffic delays and the associated delay savings attributable to the provided FSP service, graphically depicted in [Figure 1](#). Deterministic queuing and queuing diagrams originally discussed in the freeway operations context by Moskowitz (3) have been applied in numerous studies to analyze the incident impacts (4).

When an incident occurs, the normal freeway capacity c is reduced to a lesser capacity, c_i , for the duration of the incident, T_A . If the traffic demand on the freeway, v , is greater than the remaining capacity c_i , then a queue is formed upstream of the incident. Once the incident has cleared, after T_A minutes, the built-up queue will discharge at the capacity of the freeway, c , until the queue is dissipated. The total delay (in vehicle-hours) caused by the incident is the area of the triangle OCD in Figure 1:

$$delay = \frac{(v - c_i)(c - c_i)T_A^2}{2(c - v)}$$

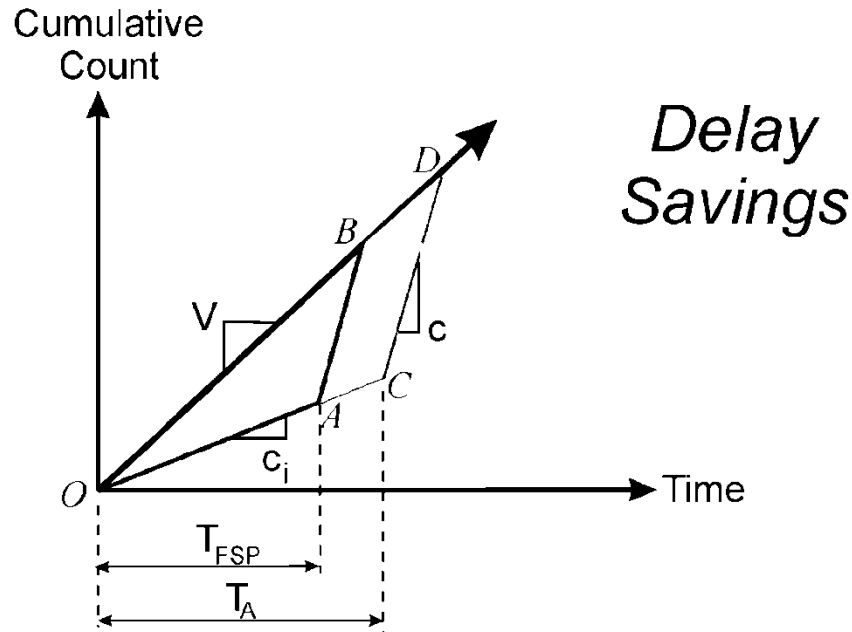


FIGURE 1 Estimation of incident delays and FSP delay savings.

The deployment of FSP results in shorter response times that reduce the incident duration (T_{FSP}) and the associated incident delay (area of triangle OAB in Figure 1). The delay savings due to FSP is the difference in delays without and with FSP service (area ABDC in Figure 1). The delay savings is attributable to FSP's faster response time. The FSP response-time reduction is the difference between the time that the FSP tow-truck arrived at the incident and the time that a tow-truck would have arrived had there been no FSP service on the beat. It is assumed that without the FSP service, stranded motorists would wait for service by a member tow company or a rotational tow arranged by CHP.

The method predicts no delays when the traffic demand v is less than the remaining capacity under incident conditions, c_i . However, in reality, there is a small amount of delay to the traffic stream because of vehicle slow-downs and rubbernecking. These small delays are ignored. The delay savings (and the benefit-to-cost ratio) depend on incident frequency and characteristics (remaining capacity and duration) and the FSP beat's operating characteristics (traffic demand and freeway capacity). The benefits are greater on heavily traveled FSP beats with a high frequency of lane-blocking incidents than on free-flowing beats with mostly shoulder breakdowns.

The methodology used to validate the FSPE model is discussed next, followed by data sources in Section 4.

METHODOLOGY FOR VALIDATING THE FSPE MODEL

In a previous research effort, a method was developed to divide the total congestion along a freeway corridor into six components indicating cause of delay: 1. incidents, 2. special events, 3. lane closures, 4. adverse weather, 5. potential reduction in delays at bottlenecks that ideal ramp

metering could achieve, and 6. excess demand (5). The Caltrans PeMS system currently hosts a fully automated two-step version of this method. The first of the two steps estimates the components of nonrecurrent congestion using statistical regression. The second method locates all bottlenecks and estimates the potential reduction in traffic delays that ideal ramp metering could achieve. The method requires input data on traffic volumes and speeds, the time and location of incidents, special events, lane closures, and adverse weather. It can readily be applied to any freeway corridor with minimal calibration.

This component of the congestion model assumes that each incident, special event, lane-closure, and adverse weather condition contributes linearly to the overall delays observed in the corridor. More complicated causality between explanatory variables, such as between the bad weather and the number of accidents, was not considered in an effort to keep the number of parameters in the model reasonable. For the components of congestion research efforts, traffic volume and speed data were obtained from the Caltrans PeMS website. Using these methods, traffic delays caused by incidents can be quantified for any freeway corridor given that adequate traffic and incident data are available for the corridor. These components of congestion techniques were used to provide empirical-based estimates of incident-induced delays that could be compared to the FSPE model's delay savings estimates.

One of the main outputs of the FSPE model is the annual delay savings, in vehicle-hours, that is attributable to the provided FSP service for a freeway corridor. The expected delay savings per FSP assist can be easily calculated using the FSPE model inputs and outputs. Likewise, the expected delay savings per minute of incident reduction can be easily estimated. For example, if the FSPE model estimated 1,600 VHT of delay savings on a beat, and the FSP tow trucks were involved with 80 assists annually, with an average incident reduction of 5 min per assist, then the delay savings per incident-minute would be $1,600 / (80 * 5) = 4.00$ vehicle-hours per incident-minute. Traditionally, the crucial challenge for the FSPE model validation efforts was to find comparable and reliable empirical delay estimates to compare to the FSPE model output.

Fortunately, a comparable measure (traffic delays per incident-minute) for a freeway corridor can be estimated using the components of congestion techniques and a combination of Caltrans PeMS and INRIX Analytics data for selected freeway corridors where FSP service is provided.

DATA SOURCES FOR FSP MODEL VALIDATION

The two primary data sources for the FSPE model validation dataset were INRIX and PeMS.

The INRIX website provides historical and real-time traffic information, travel times, and travel time information to public agencies, businesses, and individuals. To do this, INRIX collects trillions of bytes of information about roadway speeds from nearly 100 million anonymous mobile phones, trucks, delivery vans, and other fleet vehicles equipped with GPS locator devices. The data are processed in real time, creating traffic speed information for major freeways, highways, and arterials across North America, as well as much of Europe, South America, and Africa. INRIX Analytics and INRIX User Delay Cost Analysis modules were used to provide traffic delay (congestion) and corridor travel time measures for preselected freeway corridors (i.e., FSP beats).

PeMS collects data in real time from over 39,000 individual detectors spanning the freeway system across all major metropolitan areas in the State of California. PeMS is also an archived data user service that provides over 10 years of data for historical analysis. It integrates a wide variety of information from Caltrans and other local agency systems including

- Traffic detectors,
- Incidents,
- Lane closures,
- Toll tags,
- Census traffic counts,
- Vehicle classification,
- Weight-in-motion, and
- Roadway inventory.

The Caltrans PeMS website was used to provide stationary point traffic volume and delay data (mainly from freeway loops) for the set of preselected FSP beats. The Caltrans PeMS website also collects and reports CHP-reported freeway incident data.

The minimum data required to produce an estimate of expected (average) traffic delays per incident are traffic incident data and traffic delay data. PeMS was used to provide the incident data. Both INRIX and PeMS calculate and report traffic delays. This led to questions about how INRIX and PeMS estimate their traffic delays, how well the two delay estimates compare, and which estimate was the most reliable.

During the data preparation and analysis, PeMS-reported delays and INRIX-reported delays were compared to see how closely they agreed with regard to a common corridor, time period, and level of aggregation. **Figure 2** shows a scatter plot comparing the PeMS corridor reported daily traffic delays with the INRIX Analytics reported daily traffic delays for State Route 24 in California's East Bay area for the July 1, 2012, through June 30, 2013, time period. Both sets of daily traffic delays were estimated using a threshold free-flow speed of 60 mph.

The strength of the correlations between the traffic delays and the freeway incidents was used to help determine whether the INRIX or PeMS estimated delays were better suited for the components of congestion analysis techniques. It is well known that positive correlations exist between traffic delays and incident rates and that errors (e.g., measurement, estimation errors) in the traffic delay estimates and incident data only serve to deteriorate the strength of these correlations.

Table 1 shows the correlation coefficients between daily freeway traffic delays (PeMS and INRIX) and daily freeway incidents for along State Route 24 in California's East Bay area. Freeway collisions are one category of incidents in the CHP incidents database. As a sensitivity test, the correlation analysis was repeated using daily collisions (shown in Table 1).

Using the SR-24 dataset, the INRIX correlations were slightly stronger than the PeMS correlations. This finding held for incidents and collisions. The average functional detector spacing was about 1.7 mi per detector station along the SR-24 corridor.

Table 2 shows the correlation coefficients between daily freeway traffic delays and daily freeway incidents for FSP Beat 12, an 8.4-mi-stretch of I-80 in the East Bay. The average functional detector spacing was about 0.5 mi for the FSP Beat 12 corridor.

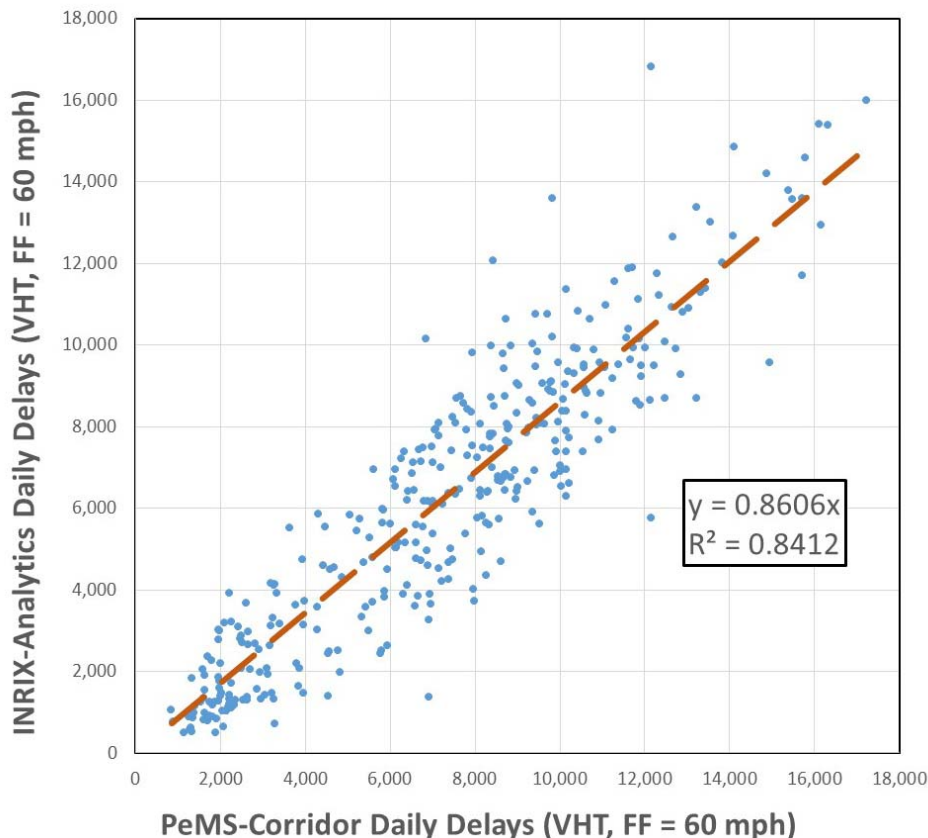


FIGURE 2 Daily traffic delays: PeMS corridor and INRIX-Analytics (State Route 24; July 1, 2012, through June 30, 2013).

**TABLE 1 Correlations Between Traffic Delay Data and Incident Data
FSP Beat 1: State Route 24; Fiscal Year 2012–2013**

Variable Sets	Incident Correlation Coefficient	Collision Correlation Coefficient
INRIX delays \leftrightarrow CHP incidents	0.4018	0.2503
PeMS delays \leftrightarrow CHP incidents	0.3454	0.2180

**TABLE 2 Correlations Between Traffic Delay Data and Incident Data
FSP Beat 12: Interstate 80; Fiscal Year 2012–2013**

Variable Sets	Incident Correlation Coefficient	Collision Correlation Coefficient
INRIX delays \leftrightarrow CHP incidents	0.3649	0.3101
PeMS delays \leftrightarrow CHP incidents	0.3573	0.3220

The PeMS (stationary detector) data correlates about as closely to the CHP incident data as did the INRIX (probe vehicle) data along the FSP Beat 12 corridor with its closely spaced PeMS detector stations.

Previous work has cited that the reliability of PeMS reported traffic data is dependent on, among other factors, a corridor's detector station spacing. For example, in an *Evaluation of PeMS to Improve the Congestion Monitoring Program* report, researchers found, "Accuracy in PeMS-based congestion estimates requires a detector spacing of less than 0.5 miles," (6). Another UC Berkeley study concluded:

Spacing loop detectors less than an average of 0.83 miles apart (i.e., using data from more than eight inductive loop detector stations along the stretch of roadway under study) did not provide extra benefit in the travel time estimation. The error remains constant between 6–13% depending on the time of day, regardless of the added loop detector stations. (7)

The reliability of the results from regression analysis, like that in the components of congestion techniques, depends heavily on the reliability of the dataset used to feed the regression analysis. As such, components of congestion analysis have typically been employed on corridors with relatively closely spaced detector stations. However, FSP service is provided on several freeway corridors with sparse or no PeMS coverage. So a validation method and data sources were sought that could be applied to FSP beats regardless of the level of PeMS coverage throughout the corridors where FSP service is provided.

Previous work has shown that PeMS detector station spacing has a direct impact on the accuracy of reported travel times and delays. What was not shown was whether PeMS detector station spacing affects the accuracy of some performance measures more than others. For example, widely spaced loops might do a better job of reliably reporting VMT than delays or vice versa.

Sensitivity testing was done on corridors with a high density of functional PeMS detector stations to gain insights into how the PeMS spacing affected key performance measures for FSP monitoring and FSPE model validation. Figure 3 shows the correlation coefficients for PeMS reported traffic volumes as a function of distance between detector stations along a corridor. It is a measure of how well traffic volumes can be approximated from measured volumes as distance from the point of measurement increases. Figure 3 also shows the same for approximating traffic delays from upstream measured delays.

As PeMS detector spacing increases, the ability to approximate traffic volumes deteriorates slowly as compared to the ability to approximate traffic delays. This implies that for corridors with widely spaced PeMS Stations, INRIX delay data might provide better delay estimates than PeMS. This is consistent with findings presented previously in Tables 1 and 2.

STUDY FINDINGS, FSPE MODEL VALIDATION RESULTS

Linear regression techniques were used to estimate the expected (average) traffic delays attributable to freeway collisions. For this FSPE model validation effort, one year's worth of CHP, PeMS, and INRIX data were compiled: July 1, 2012, through June 30, 2013, for FSP beats (corridors) listed in Table 3.

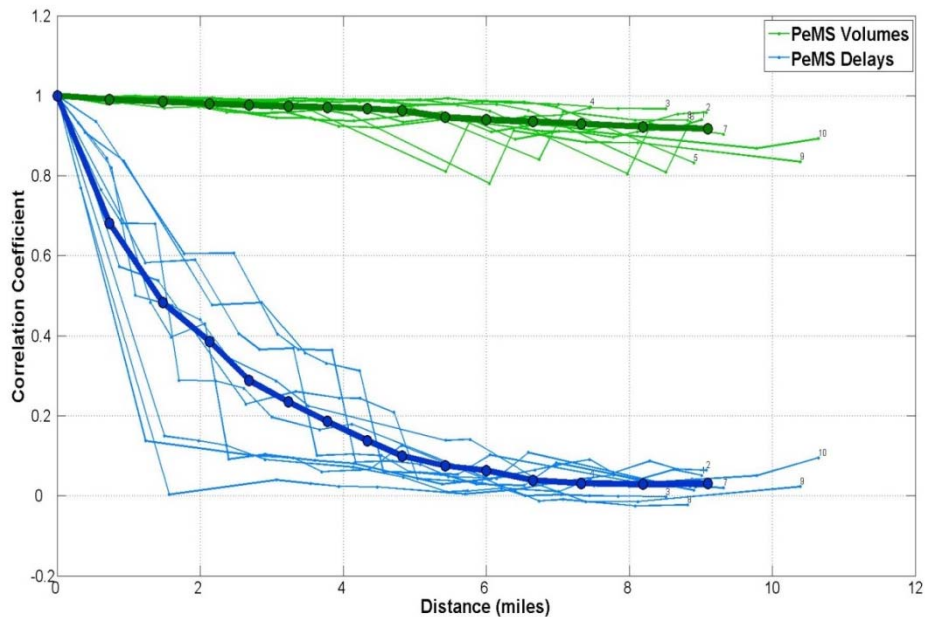


FIGURE 3 Correlations for traffic delays at PeMS stations, I-15 NB San Diego, California.

TABLE 3 Beats Selected for FSPE Model Validation

FSP Beat	County	Freeway	Beat Limits	One-Way Beat Length (mi)	Weekday FSP Trucks
1	ALA	24	I-580 to Contra Costa Co. line	4.39	2
	CC		Contra Costa Co. line to Oak Hill Rd.	6.25	
	ALA	980	I-580 to I-880	2.03	
		880	7th St. to Jackson St.	2.04	
12	CC	80	San Pablo Dam Rd. to Cummings Skyway	8.39	2
16	SCL	17	Junction SR-9 to Summit Rd.	7.07	1
18	SCL	880	Junction SR-237 to Alameda Co. line	2.08	2
	ALA		SCL Co. line to Mowry Ave.	7.18	
22	ALA	580	Santa Rita to Grant Line Rd.	16.48	3
29	SOL	80	Magazine St. to Abernathy Rd.	14.04	2
34	SOL	80	Abernathy Rd. to Vaca Valley Rd.	12.54	2
37	SOL	80	Junction I-505 to Richards Blvd.	16.40	2

On weekdays in the Bay Area, FSP provides service from 6:00 to 10:00 a.m. and from 3:00 to 7:00 p.m., although some beats operate from 5:30 to 9:30 a.m. in the mornings, and Friday afternoon shifts might vary on some beats. Sunday (weekend) FSP service is generally provided with one truck operating from 12:30 to 7:00 p.m.

Since the PeMS and INRIX data sources provided traffic delay measures, PeMS from stationary source (loop) detectors and INRIX from probe vehicles, three different measures of vehicular delays were used in the regression analysis to gain insights on how the chosen delay data source affected the regression model goodness of fit and parameter estimates:

1. PeMS traffic delays: from stationary detectors (e.g., loops),
2. INRIX traffic delays: from a relatively large sample of probe vehicles, and
3. Composite of INRIX (per-vehicle) delays and PeMS traffic volumes.

Table 4 displays the FSPE model validation results.

The findings in Table 4 are consistent with the findings presented in Tables 1 and 2. For Beat 1, the model using “INRIX (per-vehicle) Delays & PeMS Traffic Volumes” performed best. For Beat 12, the model using “PeMS Traffic Delays” performed best. Beat 12 on I-80 has an average PeMS detector spacing of 0.5 mi/station, whereas Beat 1 on SR-24 has an average of 1.7 mi between PeMS detector stations. These findings are consistent with the postulate that higher density of detector stations provides higher reliability in estimated performance measures.

The root mean square error term was calculated from the empirical and FSPE model estimated delay values, and the average regression model *F*-statistic was calculated for the FSPE model validation dataset (see **Table 5**). Overall, the INRIX delay data provided models with better model fit statistics than those created using the PeMS delay data and better than those created using the composite PeMS volume and INRIX delays.

The overall average error term (RSME) was lowest for the PeMS delay-based regression models.

Figure 4 shows a scatter plot comparing the FSPE model estimated delay savings against the empirically estimated traffic delays.

From the trend lines shown in Figure 4, the FSPE model predicts delay savings that are a very close match to the empirically estimated traffic delays using PeMS delay data. However, the FSPE model fairly significantly underestimates delay savings when compared to the traffic delays estimated using INRIX delay data. It is not clear which delay estimation (PeMS or INRIX) is more reliable without further probing into the PeMS and INRIX data collection and delay estimation procedures, and taking into account measurement and sampling errors associated with PeMS detector spacing and INRIX probe vehicle sample size.

TABLE 4 FSP Beat Evaluation Model Validation Results

Bay Area FSP Beat	FSPE Delay Savings (per min)	Source of Traffic Delay Estimate	Regression Model Traffic Delay (per min)	Regression Model Std Err of Delay	Regression Model Lower 95% Delay	Regression Model Upper 95% Delay	Regression Model F-Statistic
Beat 1 Weekday	11.44	PeMS	8.11	1.79	4.61	11.61	20.65
		INRIX	10.68	1.84	7.08	14.27	33.81
		PeMS+INRIX	12.77	2.00	8.86	16.68	40.96
Beat 1 Weekend	0.78	PeMS	-1.73	1.27	-4.22	0.77	1.84
		INRIX	-1.33	1.44	-4.15	1.48	0.86
		PeMS+INRIX	-0.48	1.68	-3.78	2.82	0.08
Beat 12 Weekday	5.26	PeMS	4.22	0.83	2.59	5.84	25.76
		INRIX	6.67	1.17	4.39	8.96	33.34
		PeMS+INRIX	6.56	1.08	4.44	8.67	36.83
Beat 12 Weekend	0.91	PeMS	0.20	0.46	-0.70	1.10	0.19
		INRIX	0.60	0.19	0.23	0.98	10.08
		PeMS+INRIX	0.91	0.33	0.26	1.55	7.66
Beat 16 Weekday	3.49	PeMS	na	na	na	na	na
		INRIX	2.09	0.72	0.68	3.50	45.70
		PeMS+INRIX	na	na	na	na	na
Beat 16 Weekend	0.75	PeMS	na	na	na	na	na
		INRIX	0.60	1.38	-2.11	3.31	0.19
		PeMS+INRIX	na	na	na	na	na
Beat 18 Weekday	7.37	PeMS	3.66	0.83	2.04	5.27	19.63
		INRIX	5.49	1.06	3.42	7.56	26.97
		PeMS+INRIX	5.13	1.01	3.15	7.11	25.79
Beat 22 Weekday	6.43	PeMS	7.61	1.15	5.36	9.86	43.87
		INRIX	17.53	2.16	13.29	21.76	65.77
		PeMS+INRIX	17.47	1.97	13.61	21.33	78.72
Beat 22 Weekend	0.45	PeMS	6.42	1.17	4.14	8.71	30.36
		INRIX	6.50	1.87	2.83	10.17	12.22
		PeMS+INRIX	8.40	1.79	4.88	11.92	21.92
Beat 29 Weekday	5.24	PeMS	1.22	0.71	-0.17	2.62	2.97
		INRIX	4.61	0.82	3.00	6.22	31.32
		PeMS+INRIX	2.00	0.43	1.16	2.85	21.52
Beat 29 Weekend	0.67	PeMS	2.01	0.98	0.09	3.93	4.19
		INRIX	1.95	0.94	0.11	3.78	4.30
		PeMS+INRIX	2.14	1.05	0.08	4.21	4.14

(continued on next page)

TABLE 4 (continued) FSP Beat Evaluation Model Validation Results

Bay Area FSP Beat	FSPE Delay Savings (per min)	Source of Traffic Delay Estimate	Regression Model Traffic Delay (per min)	Regression Model Std Err of Delay	Regression Model Lower 95% Delay	Regression Model Upper 95% Delay	Regression Model <i>F</i> -Statistic
Beat 34 Weekday	5.58	PeMS	2.87	0.53	1.83	3.90	29.29
		INRIX	12.92	2.04	8.93	16.92	40.23
		PeMS+INRIX	1.03	0.72	-0.38	2.44	2.04
Beat 34 Weekend	0.62	PeMS	-0.42	0.67	-1.74	0.90	0.39
		INRIX	1.08	0.94	-0.77	2.94	1.32
		PeMS+INRIX	0.17	0.21	-0.24	0.57	0.66
Beat 37 Weekday	5.77	PeMS	1.16	0.42	0.33	1.98	7.60
		INRIX	9.32	1.46	6.46	12.18	40.74
		PeMS+INRIX	6.73	1.14	4.49	8.97	34.65
Beat 37 Weekend	0.69	PeMS	0.89	0.74	-0.55	2.33	1.46
		INRIX	1.61	1.32	-0.98	4.21	1.49
		PeMS+INRIX	2.25	1.45	-0.60	5.09	2.40

TABLE 5 Overall Regression Model Goodness of Fit Statistics

Average FSPE Model Delay Savings (per min)	Source of Traffic Delay Estimate	Average Regression Model Delays (per min)	Average Regression Model <i>F</i> -Statistic	FSPE Versus Regression Model RSME
3.40	PeMS	2.79	14.48	2.72
	INRIX	5.35	23.22	4.42
	PeMS+INRIX	5.01	21.34	4.48

CONCLUSIONS

Overall, the FSP beat evaluation model replicated delay savings estimates that were in the range of the empirically estimated traffic delays. However, there is some evidence that the delay savings component of the FSPE model might be underestimating overall delay savings.

There are sets of plausible factors that might be contributing to the FSPE model’s underestimate of delay savings, for example, if the FSPE model’s default capacities are higher than real-world freeway capacities, or if the deterministic queueing methods used in the FSPE model tend to underestimate delays on congested freeway corridors by failing to capture the nonlinear nature of queueing, delays, and delay savings.

It should be noted that for this model validation effort, the FSPE model’s default capacity and other model parameters were used without calibration or adjustments. No fine-tuning was done to the FSPE model’s parameters or inputs to improve how well the FSPE model’s delay savings compared to the empirically estimated delay estimates. Using the default capacity, like



FIGURE 4 Empirically estimated traffic delays versus FSPE model predicted delay savings.

was done for this validation effort, might underestimate congestion for highly constrained merge, diverge, or weaving sections. Likewise using the default capacity might result in underestimated FSP delay savings for freeway segments with hills, tight curves, narrow lanes, and other geometric conditions that impact the carrying capacity of freeways. Model users do not take adequate care in ensuring the traffic volumes and other inputs are reasonable and in selecting capacity estimates that are representative of freeway geometry and traffic conditions.

The key to using any model, the FSP beat evaluation model included, is to understand the model's strengths and limitations, take care in preparing the model inputs, and perform reality checks on the model's outputs to ensure consistency with observed real-world traffic behavior.

These research efforts validated one of the components of the FSP beat evaluation (FSPE) model: the FSPE model's deterministic queueing techniques that estimate delay savings. Next steps with respect to FSPE model improvement include exploring whether using stochastic queueing methods instead of deterministic queueing methods would help to improve the FSPE model's ability to replicate real-world traffic delays and FSP delay savings.

The research support efforts for the FSP program generally focus on providing information to enable performance-based decision making. With this, two plausible and useful work efforts might be to

- Use the INRIX-Analytics datasets to provide calibration targets for the FSPE model; or, perhaps a method could be developed to directly incorporate the INRIX estimated delays into the FSP beat performance evaluation process. This would be especially helpful for freeway corridors (i.e., FSP beats) with limited or no PeMS coverage.
- Compile annual estimates for VMT, VHT, and freeway incidents for the complete set of California's FSP beats. Compare the level of FSP service provided on each beat against the beat's empirical VMT, VHT, and incident totals as a performance measure to gauge "How

closely does the allocation of FSP resources match demand for freeway incident management services?”

- Perform a before-and-after study on a freeway corridor, directly measuring and taking a detailed look at the overall and incident-induced traffic delays along a freeway corridor with FSP service on the corridor and without FSP service on the corridor.

REFERENCES

1. Skabardonis, A., H. Noeimi, K. Petty, D. Rydzewski, P. Varaiya, and H. Al-deek. *Freeway Service Patrol Evaluation*. PATH Research Report UCB-ITS-PRR-95-5, Institute of Transportation Studies, University of California, Berkeley, 1995.
2. Skabardonis, A., K. Petty, P. Varaiya, and R. Bertini. *Los Angeles: Freeway Service Patrol Evaluation*. PATH Report UCB-ITS-PRR-98-31, Institute of Transportation Studies, University of California, Berkeley, 1998.
3. Moskowitz, K., and L. Newman. Notes on Freeway Capacity. In *Highway Research Record 27*, HRB, National Research Council, Washington, D.C., 1963.
4. Urbanek, G. L., and R. W. Rodgers. *Alternative Surveillance Concepts and Methods for Freeway Incident Management*. Report RD-77-58/63. FHWA, U.S. Department of Transportation, 1978.
5. Kwon, J., M. Mauch, and P. Varaiya. Components of Congestion: Delay from Incidents, Special Events, Lane Closures, Weather, Potential Ramp Metering Gain, and Excess Demand. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1959*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 84–91.
6. Kwon, J., B. McCullough, K. Petty, and P. Varaiya. *Evaluation of PeMS to Improve the Congestion Monitoring Program*. California PATH Research Report UCB-ITS-PRR-2007-6, University of California, Berkeley, 2007.
7. Bayen, A., M. Sharafsaleh, and A. Patire. *Hybrid Traffic Data Collection Roadmap: Objectives and Methods*. PATH Research Report UCB-ITS_PRR-2013-2, Final Report for Task Order 2, University of California, Berkeley, 2013.

CALIBRATION

Calibrating Multilane First-Order Traffic Flow Model with Endogenous Representation of Lane-Flow Equilibrium

YASUHIRO SHIOMI

TATSUYA KOZONO

Ritsumeikan University, Japan

This study develops a multilane first-order traffic flow model for freeway networks. In the model, lane changes are considered as a stochastic behavior in that an individual driver decreases his or her disutility or cost and are represented as dynamics toward the equilibrium of lane-flow distribution along with longitudinal traffic dynamics. The proposed method can be differentiated from the previous studies in the following points: 1. The motivation of changing the lane is explicitly considered, and it is treated as utility defined by the current macroscopic traffic state; 2. A whole process of lane-changing is computed by macroscopic manner, that is, the extension of kinematic wave theory employing the information technology (IT) principle; and 3. In the model framework, the lane-flow equilibrium curve will be endogenously generated as a result of self-motivated lane changes. In addition, the parsimony representation enables the parameter calibration by using the data collected from the conventional loop detectors. As a result of the parameter calibration using the data collected at four different sites of Chugoku Expressway in Japan, including sag bottleneck, it is revealed that 1. The proposed method can represent the lane-flow distribution of any observation sites with high accuracy, and 2. The estimated parameters can reasonably explain the multilane traffic dynamics and the bottleneck phenomena on uphill of sag section.

INTRODUCTION

It is well known that under the condition of high traffic volume lane-flow distribution becomes unbalanced; more traffic tends to use a median lane rather than a middle and outer lane, which causes the deterioration of traffic capacity at bottleneck sections (1, 2, 3). As intensive development of intelligent transportation systems (ITS), active and dynamic lane management has been practically implemented. By employing the technology of ITS, balancing lane-flow distribution is one of the feasible solutions to increase the throughput of bottleneck flow (3). Besides the unbalanced lane usage, lane traffic management and control should be considered as one of the solutions to improve the efficiency and safety in case of lane regulation under road works or incidents and at the merging, diverging, and weaving sections. For traffic management to be effective, it is needless to say that a model-based decision support system consisting of traffic state estimation, traffic state prediction, and optimization and traffic control measures is essential as mentioned in Yuan et al. (4). However, because of the lack of a method for computing multilane traffic flow, including lane-change dynamics, a model-based decision support system enabling lane-based traffic management to be considered has not been realized.

This paper develops the multilane first order traffic flow model, which depicts the dynamics of lane-changing. In the model, it is assumed that each vehicle changes the lane to

improve its utility or decrease its disutility and also that the equilibrium of lane-flow distribution is achieved as the condition of stochastic user equilibrium (SUE), where all drivers believe that they cannot improve their utility by changing the lanes anymore. In the model, lane-changes are represented as the dynamics toward lane-flow equilibrium. The utility function for a vehicle to choose each lane is defined by only two parameters on the basis of the investigation about lane-changing behaviour done by Knoop et al. (5) and Shiomi et al. (6): one is a constant value implying cost breaking the keep-left (or right) rule, and the other one is the average speed depending on the fundamental diagram and the density of the lane. Such parsimony representation enables online calibration by using the real-time data from conventional loop detectors. To compute the possible solution of multilane traffic under the conservation law of traffic volume, the IT principle (7) is applied. Then, in this paper, the parameters reproducing the lane-flow distribution are estimated on the basis of the data collected by the conventional loop detectors. Based on the estimation results, the cross-sectional characteristics of lane-flow distribution at sag section are discussed.

This paper is organized as follows. In Section 2, state-of-the-art of modelling multilane traffic is described. In Section 3, the concept of lane-change dynamics and the mathematical representation of lane-flow equilibrium are described. In Section 4, the computation methods of multilane traffic flow employing IT principle are overviewed. In Section 5, the parameter calibration method employing the extended quasi-Newtonian approach is explained, and then in Section 6, the application results and discussion are described. Finally, the conclusion mentions the contribution of the paper and recommends future works.

STATE OF THE ART

Considerable scientific attention has been paid on the topic of lane-change behaviour and multilane flow modelling during the last two decades. Because lane-change is individual vehicle driving behavior, that is, whether a vehicle changes its lane totally depends on the decision making that the subject vehicle takes and the situation that the subject vehicle is in, it has been the most straightforward way to apply microscopic modelling (8–12). This approach can consider various conditions and variables that may cause making a decision to change lanes. However, because of the computational tasks and complicated model framework, it is not appropriate to apply for online and networkwide freeway traffic evaluation. The other approach is mesoscopic modelling (13, 14). In the approach, the gas-kinetic model is applied to depict longitudinal multilane traffic dynamics and lateral movement as well. In Shvestsov and Helbing (13), the proportion of lane changers is exogenously given according with the density. The motivations behind the lane change behavior are not appropriately considered. In Hoogendoorn and Bovy (14), the probability of a vehicle changing the lane is estimated by applying discrete choice theory. In this case, however, it is required to calibrate various parameters, so that more precise data is required than conventional loop detectors. Also, it is difficult to employ online and dynamic traffic estimation based on the real-time data collection.

From the macroscopic approach, Daganzo (15, 16) investigated the traffic phenomena on a multilane freeway, and proposed a traffic-flow theory based on the kinematic wave model, in which it is assumed that there are two types of vehicles: slugs, which have lower desired speed and drive on an outer lane, and rabbits, which have higher desired speed and drive on both outer and inside lanes depending on traffic conditions. It was proven that the slugs and rabbits theory

could explain the various traffic phenomena. However, the computational method on the basis of this theory to depict multilane traffic has not been developed. Laval and Daganzo (7) proposed a method to computing multilane traffic flow on the basis of kinematic wave theory and developed a model to depict the influence of the lane changers to the traffic flow. This study employed the hybrid approach in which lane changers are computed as particles and considered as moving bottlenecks. It is assumed that the number of lane-change vehicles is proportional to the differences of traveling speed among lanes. However, it is apparent that this assumption would not represent the lane-flow equilibrium curve appropriately. Besides, the hybrid approach combining microscopic and macroscopic model, which can be seen in Hong et al. (17) and Okaue and Okushima (18), is not feasible in the model-based decision support system. Tang et al. (19), Jin (20), and Jin (21) developed macroscopic models depicting lane-change traffic that considered the disturbances to the traffic flow caused by lane changes. However, the representation of the models is not in lane-specific manner.

This contribution can be differentiated from the previous studies in the following points:

1. The motivation of changing the lane is explicitly considered, and it is treated as utility defined by current macroscopic traffic state;
2. A whole process of lane-changing is computed by macroscopic manner, that is, the extension of kinematic wave theory employing the IT principle; and
3. In the model framework, the lane-flow equilibrium curve will be endogenously generated as a result of self-motivated lane changes.

The proposed model represents a lane-specific traffic dynamics with parsimony manner. Thus, it is expected that it has high feasibility for online and lane-specific traffic state estimation, prediction, and evaluation of a dynamic lane-control scheme.

MODELING LANE CHANGE AND LANE-FLOW EQUILIBRIUM

Assumptions of Motivations Behind Lane Changes

This study develops the model depicting lane-changing dynamics and lane-flow equilibrium at a freeway section without any merging and diverging, where all vehicles change the lane to improve their driving circumstances. Namely, mandatory lane-changes heading to off-ramp or coming from on-ramp are not considered.

It is well known that on such sections the specific macroscopic relationship between the total density and the fraction of the lane flow as shown in Figure 1 can be observed. On two-lane sections, more traffic tends to use on the outside lane when traffic density is not so large, while under the presence of higher traffic density than approximately 30 (veh/km/2lanes), more traffic drives on the median lane rather than the other lane. As the density increases, the gap of fraction becomes insignificant. On three-lane sections, it is more complicated than on two-lane sections. First, the fraction of the outside lane is more than the other lanes in the presence of less traffic density than 20 (veh/km), and then traffic on the center lane becomes dominant. In the higher traffic density than 50 (veh/km), the fraction of the median lane becomes the largest, and finally

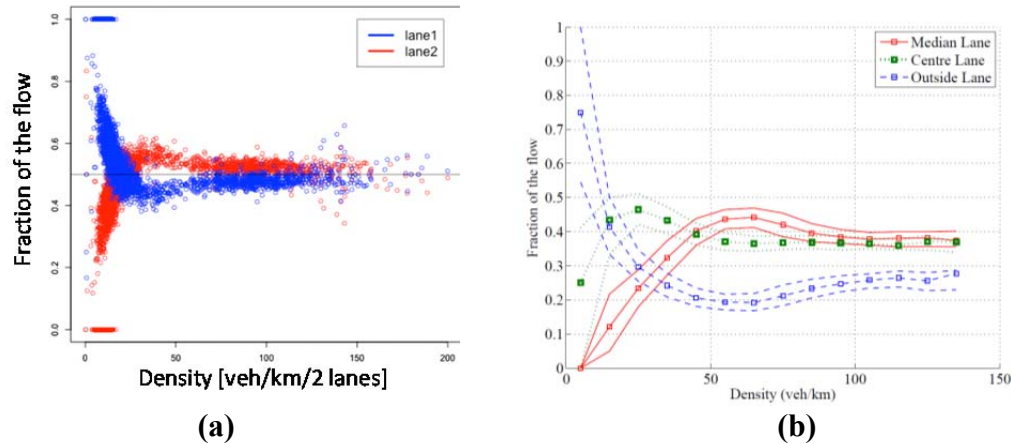


FIGURE 1 Observation of lane-flow distributions in freeways:
(a) example of two-lane section and (b) example of three-lane section (1)

the gap of fractions among the lanes gets insignificant. This tendency is not special to the observation site shown in Figure 1 but can be observed generally all over the world.

With regard to the mechanism of the lane-flow distribution and its equilibrium condition, Wu (2) theoretically revealed that the equilibrium curve was achieved as a result of balance between the lane-change demand, that is, the proportion of the following vehicles that are forced to drive less than their desired speed and the proportion of the available gap in the adjacent lanes. From the empirical aspects, Knoop et al. (5) investigated the relationship between the number of lane-changes and the density of both original and adjacent lanes. It revealed that in the free-flow condition the number of lane changes per traffic volume from an outer lane to a median lane and vice versa is not negatively proportional to the density of the adjacent lane if the density of the original lane is the same level. This fact implies that lane-change behaviors are not fully explained only by the gap acceptance, that is, the proportion of the available gap. According to Shiomi et al. (6), which investigated lane-change behaviors on three-lane sections by applying a discrete choice model, in the outside and middle lanes, vehicles tend to remain in the original lane, whereas in the median lane, vehicles tend to change lanes to either the middle or outer lanes. This fact indicates that drivers basically compliant with the keep-left (in case of Japan) rule, which also motivates drivers to change the lane or remain on the same lane.

Thus, in this study, it is assumed that

1. Drivers are motivated to change the lane to increase the driving speed, though it depends on their desired speed; that is, a driver with high desired speed would change the lane and one with low desired speed would not try that.
2. Basically, drivers would follow the keep-left rule. That is, if the traffic state is the same among lanes, a driver would choose the outside lane.
3. The demand of lane changes is censored because of the limitation of the available gap on the target lane, which is mutually related to the available capacity of the target lane.

Then, under these assumptions, a first-order traffic flow model to depict multilane traffic dynamics is developed.

Definition of Utility Function and Mathematical Expression of Equilibrium State

Suppose a fundamental diagram is defined lane by lane, and the average speed of lane l , v_l , is given as

$$v_l = f_l(k_l)$$

where $f_l(\cdot)$ is a fundamental relationship and k_l is the density on lane l , respectively. A driver would choose a lane that gives him or her more utility or less disutility. As mentioned in the previous section, a driver would be motivated to change the lane to increase his or her driving speed or to follow the keep-left rule. Thus, the cost of a vehicle n to drive on lane l , $c_{nl}(k_l)$, is defined as a monotonically increasing function against the density:

$$c_{nl}(k_l) = \alpha_l + \beta_l \cdot \{f_l(k_l)\}^{-1} + \varepsilon \quad (1)$$

where $k_l(t, x)$ is the density on lane l at time-space point (t, x) , α_l is the disutility to violate the keep-left rule, β_l shows sensitivity to the travel time of a unit of distance, and ε is an error term following Weibull distribution, $W(0, \theta)$, implying the heterogeneity of desired speed and recognition error. Assuming traffic flow is composed of homogenous vehicles in terms of their fundamental diagram and the structure of the cost function, the probability that a vehicle chooses lane l according with the current traffic situation at time t is written as

$$p_l(K) = \frac{\exp[-\theta \cdot c_l(k_l)]}{\sum_k \exp[-\theta \cdot c_k(k_k)]}$$

where $K(t, x)$ is the total density and written as

$$K(t, x) = \sum_l k_l(t, x)$$

Note that an index n is omitted because of the clear representation. The lane-flow equilibrium condition means the state where each driver believes that he or she can no longer decrease the driving cost by changing the lane, or the situation where even if some vehicle change their lanes, other vehicles would compensate for the change of lane traffic flow by changing the lane immediately and as a result lane flow distribution becomes stable. The equilibrium state is indicated by Equation 2.

$$\begin{aligned} p_l^*(K) &= \frac{\exp[-\theta \cdot c_l(k_l^*)]}{\sum_k \exp[-\theta \cdot c_k(k_k^*)]} \\ &= \frac{k_l^*}{K} \end{aligned} \quad (2)$$

where $*$ is the symbol indicating the equilibrium condition.

Expression of Lane-Change Dynamics

The equilibrium condition expressed by Equation 2 is equivalent to the solution of the optimization problem as

$$\begin{aligned} \min Z(\mathbf{k}) &= \sum_k \int_0^{k_k} c_k(\omega) d\omega + \frac{1}{\theta} \sum_k k_k \ln \frac{k_k}{K} \\ \text{subject to} & \\ K &= \sum_k k_k \\ k_k &\geq 0 \end{aligned} \quad [\text{P-1}]$$

because the equilibrium condition can be considered as SUE condition, and the cost function is monotonic increasing function with regard to the density. To solve the problem [P-1], the objective function is partially linearized as

$$\begin{aligned} \min Z(\mathbf{y}) &= \sum_k y_k c_k(z_k) + \frac{1}{\theta} \sum_k y_k \ln \frac{y_k}{Y} \\ \text{subject to} & \\ Y &= \sum_k y_k \\ y_k &\geq 0 \end{aligned} \quad [\text{P-2}]$$

where $\mathbf{z} = \{z_k\}$ is a vector of the density on lane k at (t, x) . Then, the solution vector \mathbf{y}^* is given by calculating KKT condition as

$$y_k^* = Y \cdot \frac{\exp[-\theta \cdot c_k(z_k)]}{\sum_j \exp[-\theta \cdot c_j(z_j)]} \quad (3)$$

It is mathematically proven that the operation

$$\mathbf{z}(t + \Delta t, x) = \mathbf{z}(t, x) + (1/\tau) \{\mathbf{y}^* - \mathbf{z}(t, x)\} \quad (4)$$

gives a better solution of [P-1] than $\mathbf{z}(t, x)$, where $\tau > 1$ (22). This result implies that when the cost of each lane is defined as a monotonic increasing function with regard to the density, the lane-flow distribution gradually approaches the equilibrium condition as vehicles repeatedly change lanes in an ad hoc manner following the choice probability. In this model, the process toward the equilibrium represents the dynamics of lane change. In Equation 4 a dynamic parameter, τ , is used. It can be interpreted as the same line as τ of Laval and Daganzo (7). Namely, τ indicates the number of time steps a driver takes to decide and execute a lane change. It relates to the gap availability in the target lane. The higher the density of the target lane is, the longer time to find an available gap takes. Thus, τ is considered as such a parameter that relates to traffic condition, and becomes larger when the density is higher.

MULTILANE FIRST-ORDER TRAFFIC FLOW MODEL

Framework of Multilane LWR Model

In this study, the multilane Lighthill-Whitham-Richards (LWR) model developed by Laval and Daganzo (7) is applied with some modification to compute traffic flow on multilane with lane changes. The conservation law of multilane traffic is written as

$$\frac{\partial K_l(t, x)}{\partial t} + \frac{\partial Q_l(t, x)}{\partial x} = \phi_l, \quad l = 1, 2, \dots, n \quad (5)$$

where $K_l(t, x)$ and $Q_l(t, x)$ indicate the density and traffic flow on lane l in position x at time t , respectively. The nonhomogeneous term, ϕ_l , in Equation 5 shows the balance caused by the lane-change vehicles. Thus, this term can be rewritten as

$$\phi_l = \sum_{l' \neq l} \phi_{l' \rightarrow l} - \sum_{l' \neq l} \phi_{l \rightarrow l'}$$

where $\phi_{l' \rightarrow l}$ means the number of vehicles coming from the other lane (l') to the target lane (l).

As mentioned above, it is assumed that fundamental diagram is defined lane by lane. Note that lane change is caused by the differences in the cost among lanes even in the free flow condition. Thus, the equation of fundamental diagram proposed by van Lint et al. (23) is used. It is shown as follows:

$$V_l(t, x) = f_l(K_l) = \begin{cases} v_{fl} - K_l \cdot \frac{v_{fl} - v_{cl}}{k_{cl}} & \text{if } 0 \leq K_l \leq k_{cl} \\ \frac{v_{cl} \cdot k_{cl}}{K_l} \cdot \left(1 - \frac{K_l - k_{cl}}{k_{jl} - k_{cl}} \right) & \text{otherwise} \end{cases} \quad (6)$$

where v_{fl} , v_{cl} , k_{cl} and k_{jl} show the free flow speed (km/h), critical speed (km/h), critical density (veh/km), and jam density (veh/km) on lane l , respectively. The example of the fundamental diagram following to Equation 6 is exhibited as [Figure 2](#).

Extension to Multilane Traffic Flow

Godunov Scheme for Multilane Section

To compute the traffic dynamics following the conservation law in Equation 5, the Godunov scheme is applied. Then, Equation 5 is discretized as

$$\frac{K_{t+1,il} - K_{til}}{\Delta t} + \frac{Q_{til} - Q_{t,i-1,l}}{\Delta x} = \sum_{l' \neq l} \phi_{i,l' \rightarrow l} - \sum_{l' \neq l} \phi_{i,l \rightarrow l'}$$

where the index t and i show time step and cell number, respectively. Following to CFL condition, the time step size Δt and cell length Δx should keep the constrained condition as follows:

$$\Delta x \geq \max_{\forall l} (v_{fl}) \cdot \Delta t$$

In the case of a single-lane section, the traffic volume transferred from the upper cell i to the downer cell $i+1$, A_{ii} , is given as

$$A_{ii} = \min(S_{ii}, R_{t,i+1}, (k_{J,i+1} - K_{t,i+1}) \cdot \Delta x)$$

where

$$S_{ii} = \begin{cases} K_{ti} \cdot V_{ti} \cdot \Delta t & \text{if } 0 \leq K_{ti} \leq k_{ci} \\ k_{ci} \cdot v_{ci} \cdot \Delta t & \text{otherwise} \end{cases}$$

$$R_{t,i+1} = \begin{cases} k_{ci+1} \cdot v_{ci+1} \cdot \Delta t & \text{if } 0 \leq K_{t,i+1} \leq k_{ci+1} \\ K_{t,i+1} \cdot V_{t,i+1} \cdot \Delta t & \text{otherwise} \end{cases}$$
(7)

In Equation 7, S_{ii} is the sending function, which means that the traffic demand from the upper cell, and $R_{t,i+1}$ is the receiving function, which means the supply volume of the downer cell. The transfer volume is limited as the minimum of the traffic demand, supply volume, and the physically acceptable number of vehicles in the downer cell. In the following section, the treatment of lane-change vehicles is given.

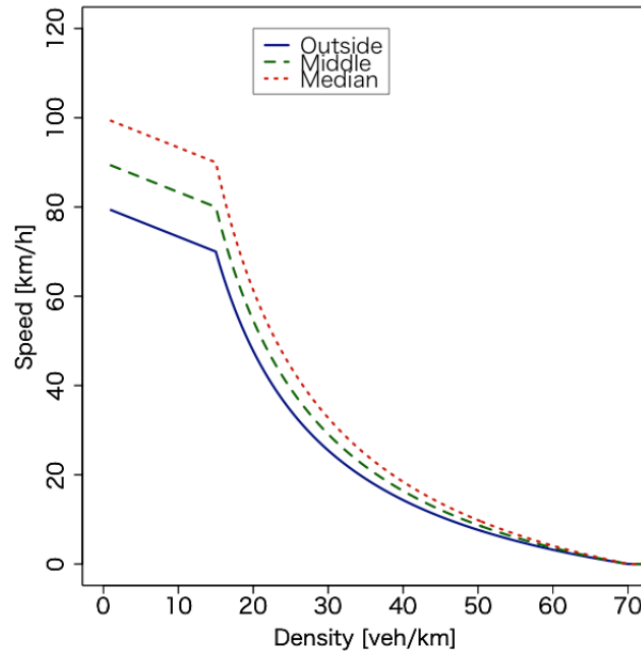


FIGURE 2 An example of fundamental diagram, where $v_f = \{80, 90, 100\}$, $v_c = \{70, 80, 90\}$, $k_c = \{15, 15, 15\}$, and $k_j = \{70, 70, 70\}$.

Definition of the Number of the Vehicles with Desire of Lane Change

As mentioned in Section 3, lane-change vehicles are generated to improve their driving cost. Given the cost function to each lane in accordance with the density of each cell as with Equation 1, the proportion of the vehicles with the desire to change the lane from l to l' , in time and place (t, x) , $p_{il,l \rightarrow l'}$, is defined as Equation 8.

$$p_{il,l \rightarrow l'} = \frac{\exp[-\theta \cdot c_{il'}(K_{il'})]}{\sum_k \exp[-\theta \cdot c_{ik}(K_{ik})]} \quad (8)$$

Then, the number of vehicles with desire to change the lane is written in accordance with the sending function S_{il} as follows:

$$L_{il \rightarrow l'} = \frac{1}{\tau_{il}} \cdot S_{il} \cdot p_{il \rightarrow l'}$$

where τ_{il} is the dynamics parameter. It is an unknown parameter depending on the traffic state. It is not able to be observed directly so that should be estimated on the basis of the longitudinal variation of lane-flow distribution. Along this line, a feedback estimation method [for example, in Wang and Papageorgiou (24)] could be applied to determine the parameter. The volume of the traffic with the desire to keep the lane is also defined as

$$M_{il} = S_{il} - \sum_{l' \in \Omega_l} L_{il \rightarrow l'}$$

where Ω shows the set of lanes that a vehicle can get to within time step Δt from the current lane l .

Computing Lane-Change Vehicles

Based on the number of the vehicles with the desire to keep the lane and the number of the vehicles with the desire to change the lane, the adjustment process to determine the transfer volume into the downer cells. The IT principle (7) is applied with partial revisions. In this study, it is assumed that there are two criteria to execute a lane change. The first criterion is whether a vehicle can find a space in the target lane, and the other criterion is whether the downstream cell on the target lane can accept the traffic coming from the upstream of the same lane and its adjacent lanes.

Let H_{ii+1l} denote the total desired number flowing into the cell $i + 1$ of lane l on time t as

$$H_{ii+1l} = M_{il} + \sum_{l' \in \Omega_l} L_{il' \rightarrow l} \quad (9)$$

For Equation 9, the first criterion is applied, that is, the desired number of lane changes is censored according to the acceptable volume on the adjacent cell on the target lane, which makes H_{ii+1l} denoted by

$$H'_{t,i+1,l} = M_{til} + \gamma_{til} \cdot \sum_{l' \in \Omega_t} L_{til' \rightarrow l}$$

where

$$\gamma_{til} = \min \left(1, \frac{\min(R_{til}, (k_{ji} - K_{ti})\Delta x)}{H_{ti+1l}} \right)$$

Then, the second criterion is applied. Let ω_{til} denotes,

$$\omega_{til} = \min \left(1, \frac{\min(R_{ti+1l}, (k_{ji+1} - K_{ti+1})\Delta x)}{H'_{ti+1l}} \right)$$

which defines the possible transfer volume with keeping the lane, q_{til} , and the possible transfer volume with lane change, $\phi_{til' \rightarrow l}$, as

$$q_{til} = \omega_{til} \cdot M_{til}$$

$$\phi_{til' \rightarrow l} = \omega_{til} \cdot \gamma_{til} \cdot L_{til' \rightarrow l}$$

respectively. Then, the actual traffic volume flowing into the downstream cell $i + 1$ on lane l , A_{ti+1l} , is written as

$$A_{ti+1l} = q_{til} + \sum_{l' \neq l} \phi_{til' \rightarrow l}$$

Finally, the density of each cell is updated every time step in accordance with

$$K_{t+1il} = K_{til} + \left\{ A_{til} - q_{til} - \sum_{l' \neq l} \phi_{til \rightarrow l'} \right\} \cdot \Delta x$$

PARAMETER CALIBRATION METHOD

The proposed model requires the following parameters:

- Fundamental diagrams: free speed, critical density, critical speed, and jam density for each lane, and
- Utility function of lane choice: disutility to violate the keep-left rule, and sensitivity to the travel time of a unit of distance for each lane.

In this section, the parameter estimation method on the basis of the data collected by the conventional loop detectors is proposed. Here, the dynamic parameter, τ in Equation 4 is excluded in the estimated parameters. It is assumed that the parameters of the utility function of lane choice are defined under given fundamental diagrams. Thus, the parameters are estimated first for fundamental diagrams, followed by those of utility function of lane choice.

Estimation Method of FD Parameters

Conventional loop detectors can directly measure the lane-based traffic volume ($Q_{obs,l}$) and space-mean speed ($V_{obs,l}$), and the density ($K_{obs,l}$) can be indirectly measured by the operation, $K_{obs,l} = Q_{obs,l}/V_{obs,l}$. Suppose Θ_l denotes the unknown parameter vector in FD on lane l , the estimation vector Θ_l^* , which minimizes the residual error between the estimations and the observations, can be found by Equation 10.

$$\Theta_l^* = \arg \min_{\Theta_l} \left[\sum_i^{N_l} \left\{ V_{obs,l}^i - f_l(K_{obs,l}^i | \Theta_l) \right\}^2 \right] \quad (10)$$

where N_l shows the number of observations on lane l . To solve Equation 10, quasi-Newton's method is applied.

Estimation Method of Lane-Choice Parameters

For the calibration of lane-flow distribution, α_l and β_l in Equation 1 are unknown parameters to be calibrated. According to Equation 2, it is obvious that only the relative difference of the cost among lanes has influence on the lane-choice behavior, so that α_1 and β_1 on the outside lane are given as 0 and 1, respectively. Then, given the total density K and the unknown parameter set Φ , the proportion of lane flow at the equilibrium condition, $\mathbf{p}_{est}(K|\Phi) = \{p_{est,l}(K|\Phi) | l = 1, 2, \dots, n\}$, where $\sum_l p_{est,l}(K|\Phi) = 1$, is obtained by simulating traffic flow on an imaginary ring road with periodic boundaries. To get the convergence results, the dynamic parameter, τ , is set as the same with the number of time steps in the simulation.

From the conventional loop detectors, the cross-sectional traffic density, K_{obs} , and the proportion of lane flow on lane l , $p_{obs,l}$, can be observed. It requires huge computational tasks to get the convergence results for all the observed density, so that the estimation of the lane-flow distribution is obtained by the following approximating method. Suppose \mathbf{x} denotes the finite-discrete point sequences shown as $\mathbf{x} = \{x_1, x_2, \dots, x_n | 0 < x_1 < x_2 < \dots < x_n < K\}$. The set of convergence solutions of lane-flow equilibrium corresponding to the densities \mathbf{x} , $\mathbf{P}_{est}(\mathbf{x}|\Phi) = \{p_{est}(x_1|\Phi), \dots, p_{est}(x_n|\Phi)\}$ under the given parameter set Φ is computed in advance. Then, the estimate of the proportion of lane flow on lane l corresponding to the observed density, K_{obs} , is obtained as a linear interpolation as follows:

$$\hat{p}_{est,l}(K_{obs}|\Phi) = \frac{p_{est,l}(x_i|\Phi) \cdot (x_{i+1} - K_{obs}) + p_{est,l}(x_{i+1}|\Phi) \cdot (K_{obs} - x_i)}{x_{i+1} - x_i} \quad (11)$$

where $x_i \leq K_{obs} < x_{i+1}$.

The parameter set Φ is found by minimizing the residual errors as follows:

$$\Phi^* = \arg \min_{\Phi} \left[\sum_i^N \sum_l^L \left\{ \hat{p}_{est,l}(K_{obs}^i|\Phi) - p_{obs,l}^i \right\}^2 \right] \quad (12)$$

where L is the number of lanes. Equation 12 can be solved by applying quasi-Newton's method.

APPLICATIONS

The proposed method is applied to the real field data. First, the details of the study site and the collected data are described, and the parameter estimation results are shown. Then, the discussion on what the results mean follows.

Study Section

The data used in the study were collected at four cross-sectional points on a three-lane section of the Chugoku Expressway in Japan, which includes the sag section, and traffic congestion often occurs behind Takaraduka–West tunnel on 20.32 KP. The geometric features of the target section and the data collection sites are illustrated in Figure 3, in which the section between 20.90 and 20.32 KP becomes the bottleneck because of sag structure. The data including 5-min space–mean speed and 5-min traffic volume were collected lane by lane at each observation point March 16–23, April 20–30, May 1–18, July 15–31, and September 1–14, 2010. The data collected under incidents or road works are excluded, so that in total the data of 30 days is used for the further analysis.

Estimation Results

FD Parameters

As the first step to calibrate the multilane traffic flow model, the parameters in FD of each lane on each observation site are estimated. The results are summarized in Figure 4 and Table 1. As seen in Figure 4 showing the comparison between the observations and the estimations, the estimates indicate a good fit to the observations except for 25.20 KP, which is located about 5 km upstream from the bottleneck. Interestingly, even in the congested flow region, the average speed in the middle lane is larger than that in the outer lanes against the same traffic density, and as the density increases the gap of the speed diminishes. This tendency is fully captured by the estimation of FD. Focusing on the traffic capacity, q_c , defined as $v_c \cdot k_c$, it appears that traffic capacity at 20.32 KP is much less than the other sites on any lane, which clearly shows 20.32 KP becomes a bottleneck.

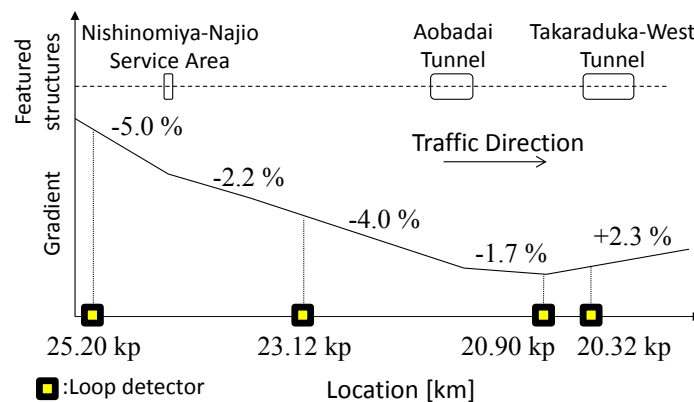


FIGURE 3 Study section.

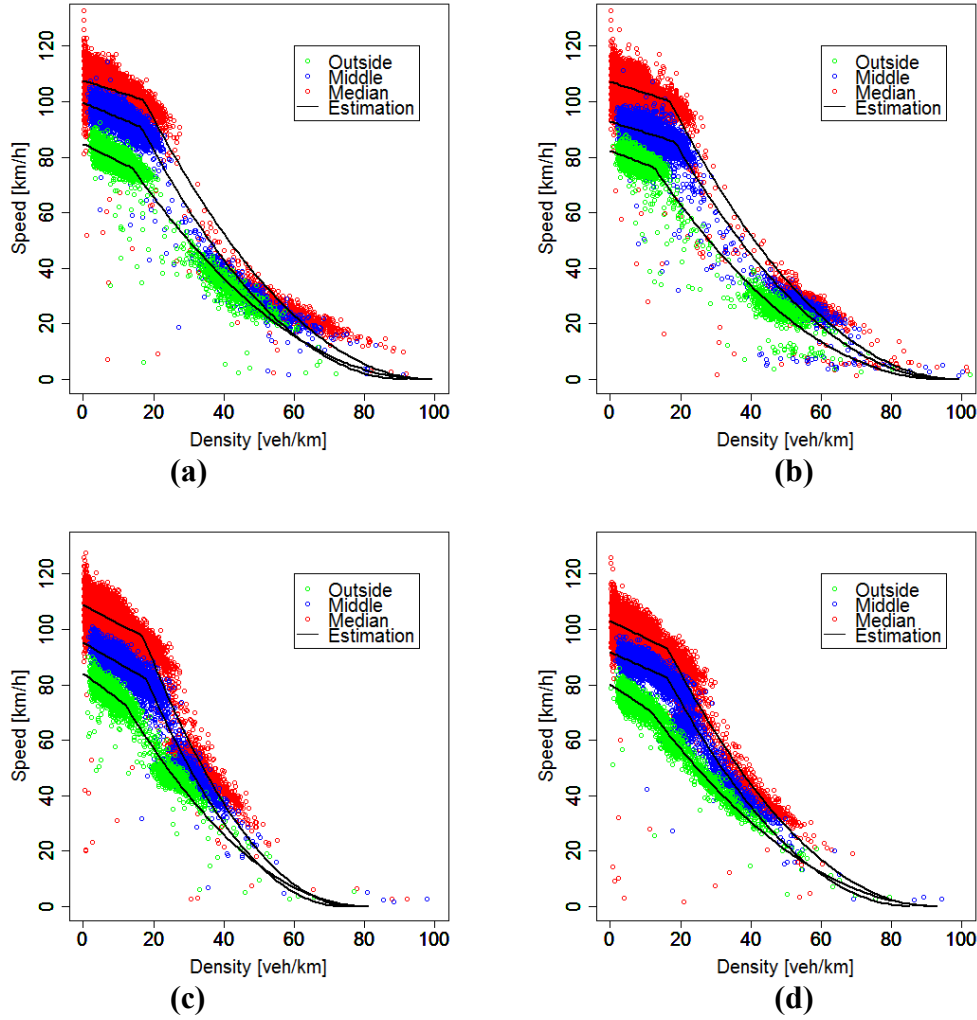


FIGURE 4 Comparisons between observations and estimations of FD: (a) 25.20 KP; (b) 23.12 KP; (c) 20.90 KP; and (d) 20.32 KP.

TABLE 1 FD Parameters

Site	Lanes	v_f (km/h)	v_c (km/h)	k_c (veh/km)	q_c (veh/h)	k_j (veh/km)
25.20 KP	Outside	84.7	75.9	14.1	1,070.2	98.2
	Median	99.6	90.7	16.6	1,505.6	91.6
	Middle	107.6	100.6	17.2	1,730.3	99.2
23.12 KP	Outside	82.3	76.2	12.5	952.5	94.7
	Median	92.8	85.6	18.4	1,575.0	97.1
	Middle	107.1	100.2	16.9	1,693.4	99.3
20.90 KP	Outside	84.1	73.0	11.9	868.7	81.5
	Median	95.2	82.1	17.9	1,469.6	73.9
	Middle	108.7	97.5	17.0	1,657.5	77.2
20.32 KP	Outside	80.2	70.6	11.5	811.9	94.6
	Median	91.7	82.5	16.3	1,344.8	86.3
	Middle	103.1	93.0	16.3	1,515.9	92.5

Parameters on Lane Choices

To estimate the parameters on the cost function for the lane choice in Equation 1, the variance parameter θ is empirically set as 1,000. The number of iterations to get the convergence solution of Problem 1 (P-1) is set as 50. The number of cells consists of the imaginary ring road is set as 3. As previously mentioned, at any observation site α_1 and β_1 is set as 0 and 1.0, respectively, to standardize the lane choice costs.

The comparisons between the observations and estimation of lane-flow distribution based on the estimated parameters are shown in Figure 5, and the estimation results are summarized in Table 2. It is interestingly noted that the range of the estimated parameters is almost the same for all observation sites. It implies the validity of the estimation results, which can be confirmed by the comparison results showing the good fits to the observations with high accuracy. Concretely, the estimation of the lane-flow distribution captures the features that when traffic density is few the fraction of outside lane is dominant, and as the traffic density increases the most dominant lanes shift to the middle lane, and finally the median lane becomes the most dominant, though this feature is slightly different among the observation sites. Thus, these results imply the validity of the approach employed in the research that the lane-flow distribution is formulated as the SUE condition, and the motivations behind the spontaneous lane changes are simply described by the compliance with the keep-left rule and the sensitivity to the increase of the travel time.

Discussions

Hereafter, the paper discusses the characteristics of lane use along the sag section on the basis of the estimated lane-choice parameters. Figure 6a shows the comparisons of the estimated parameter, α , indicating the degree of compliance with the keep-left rule. If the average speed of each lane is high, the second term of Equation 1 becomes lower, and the influence of the parameter on lane choice becomes relatively higher, that is, the parameter, α , has more influence in the light traffic situation. The larger this parameter of a lane is, the more the cost to drive on the lane is. According to Figure 6a, it is clearly shown that the value of the median lane is larger than the middle lane for any site, which implies that the median lane is basically not likely to be chosen when traffic volume is not so large. Focusing on the differences among the observation sites, it appears that on 25.20 KP, which is the beginning of the downhill section, the parameter of the middle lane is higher than that of 23.12 KP, which is the middle point of the downhill section. It means that in the light traffic the fraction of use of the outside lane at 25.20 KP is larger and gradually shifts to the middle lane toward the middle point of the downhill at 23.12 KP. At the bottom of the sag on 20.90 KP, the value of the median lane is slightly higher than 23.12 KP, while that of the middle lane is almost same. At the uphill section of 20.32 KP, the value of the middle lane gets higher as well as the median lane. These facts indicate that in the case of light traffic, on the downhill section where free flow speed tends to be high, the proportion of the middle lane becomes high, whereas on the bottom and uphill section, where free-flow speed becomes gradually lower, the proportion of lane use shifts to the outside lane. This tendency can be seen in Figure 5.

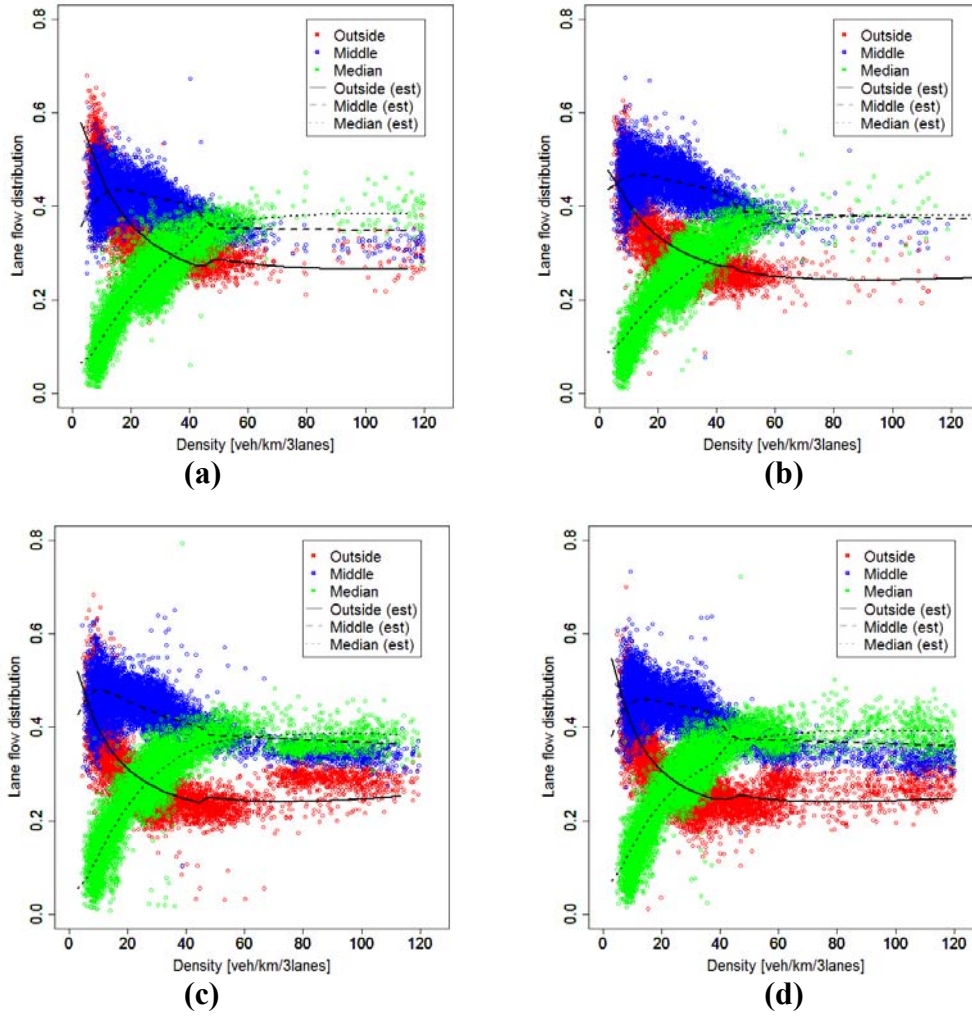


FIGURE 5 Comparisons between observations and estimations of lane-flow distribution: (a) 25.20 KP; (b) 23.12 KP; (c) 20.90 KP; and (d) 20.32 KP.

TABLE 2 Parameters on Lane Choices

Sites	25.20 KP	23.12 KP	20.90 KP	20.32 KP
α_1	0	0	0	0
α_2	0.0134	0.0120	0.0122	0.0133
α_3	0.0179	0.0181	0.0185	0.0192
β_1	1	1	1	1
β_2	0.830	0.830	0.828	0.830
β_3	0.802	0.809	0.807	0.800

Figure 6b shows the comparison results of the sensitivity parameter to travel time, β . This parameter can be interpreted as follows: The larger the parameter is, the less the lane is likely to be chosen as traffic gets congested. It should be interestingly noted that all the estimates shown in Figure 6b are less than 1.0, which is corresponding to the parameter of the outside lane. It implies that the outside lane is less likely chosen than the other lanes when traffic volume is high. Also, it can be found that the parameters of the median lanes are less than the middle lanes in any sites, which indicates that the dominant lane shifts from the outside, middle to the median as traffic volume increases. These seem to be valid results. Next, the estimation results are compared among the observation sites. As seen in the figure, the sensitivity of the median lanes on 23.12 KP (at the middle of the downhill) and 20.90 KP (at the bottom of the sag) are higher than the other sites, while no large differences in the sensitivity of the middle lanes among the observation sites are found. It means that on the section from the midpoint of the long downhill to the bottom of the sag, the traffic flow tends to use less on the median lane in comparison to the bottleneck point on 20.32 KP. In addition, the fact that the parameter on the middle lane at 20.90 KP is slightly less than the other sites implies that at the bottom of sag traffic flow tends to use more on the middle lane. The underlying mechanism inducing such results can be considered as follows: On the long downhill section, the traveling speed is high enough on average that drivers are not strongly motivated to use the median lane, rather they choose more to drive in the middle lane. When traffic flow comes to the uphill section, it slows down the driving speed on the whole, and drivers in the middle lane try to change the lane to the median to get the speed gain. As pointed out in Patire and Cassidy (25), the lateral traffic dynamics beyond lanes might cause speed disturbance, which might makes 20.32 KP a bottleneck. Although these findings are to be confirmed by the direct observations on the lane change behavior, it is revealed that the proposed method can be applied to grasp the characteristics of lane use from the macroscopic point of view.

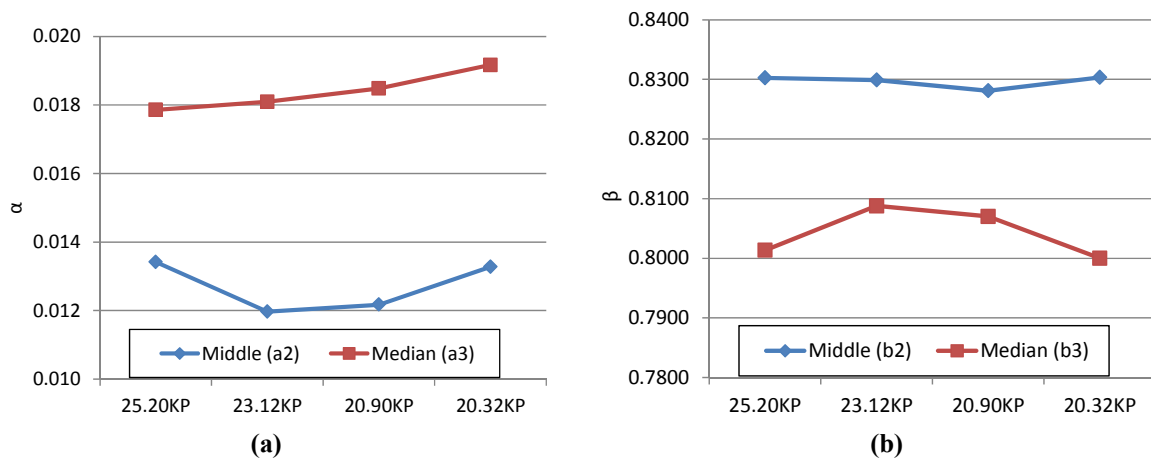


FIGURE 6 Comparisons of estimated parameter values:
(a) parameters of keep-left term; **(b)** parameters of sensitivity to travel time.

CONCLUSIONS

In this contribution, a multilane, first-order, macroscopic traffic flow model is developed. In the model, it is assumed that a driver changes lanes to improve the utility or cost of driving circumstance. The utility–cost function is composed by a constant value, indicating the cost to break the keep-left rule; a coefficient of the inverse of the speed defined by the fundamental diagram, indicating the sensitivity to the increase of the travel time; and an error term, implying the heterogeneity of drivers and the limitation of the information about the surrounding traffic situation. Then, the parameters on the cost function are successfully calibrated on the basis of the data collected from the conventional loop detectors. Also, on the basis of the estimated parameters, the multilane traffic dynamics on the section including sag are discussed. As a result of the parameter calibration using the data collected at four different sites of the Chugoku Expressway in Japan, including sag bottleneck, the following are revealed:

1. The proposed method can represent the lane flow distribution of any observation sites with high accuracy with the observations, and
2. The estimated parameters can reasonably explain the multilane traffic dynamics and the bottleneck phenomena on uphill of sag section.

In this study, the parameters of the cost function for lane choices are limitedly calibrated. Although these parameters are significant for representing the lane-flow distribution, they can limitedly capture the static characteristics of lane use. To depict the occurrence of lane changes, it is essential to calibrate the dynamics parameter, τ , which adjusts the number of lane change vehicles. For this purpose, dynamic feedback system employing Kalman filter family method would be applicable. In addition, it is recommended to upgrade the multilane traffic flow model to depict, for example, mandatory lane changes at such sections with on- and off-ramps, merging and diverging with multiclass representation.

ACKNOWLEDGMENT

This work was supported by JSPS Kakenhi Grant Number 24560640. The authors thank both of the engineers at Sumitomo System Electric Solution Co. Ltd. and West Nippon Expressway Co. Ltd., who provided the data set for the analysis.

REFERENCES

1. Knoop, V. L., A. Duret, C. Buisson, and B. van Arem. Lane Distribution of Traffic Near Merging Zones Influence of Variable Speed Limits, *Proc. IEEE Conference on Intelligent Transport Solutions*, 2010, pp. 485–490.
2. Wu, N. Equilibrium of Lane Flow-Distribution on Motorways. In *Transportation Research Record: Journal of Transportation Research Board*, No. 2278, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 31–41.
3. Xing, J., E. Muramatsu, and T. Harayama. Balance Lane Use with VMS to Mitigate Motorway Traffic Congestion, *International Journal of Intelligent Transportation Systems Research*, Vol. 12, Issue 1, 2014, pp. 26–35.

4. Yuan, Y., J. W. C. van Lint, R. E. Wilson, F. van Wageningen-Kessels, and S. P. Hoogendoorn. Real-time Lagrangian Traffic State Estimator for Freeways. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, Issue 1, 2012, pp. 59–70.
5. Knoop, V. L., S. P. Hoogendoorn, Y. Shiomi, and C. Buisson. Quantifying the Number of Lane Changes in Traffic: An Empirical Analysis. In *Transportation Research Record: Journal of Transportation Research Board*, No. 2278, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 31–41.
6. Shiomi, Y., T. Ogawa, N. Uno, H. and Shimamoto. Vehicle Re-Identification for Travel Time Measurement Based on Loop Detectors Considering Lane Changes. *TRB 92nd Annual Meeting Compendium of Papers*, 2013.
7. Laval, J., and C. F. Daganzo. Lane-Changing in Traffic Streams. *Transportation Research Part B*, Vol. 40, 2006, pp. 251–264.
8. Gipps, P. G. A Model for the Structure of Lane-Changing Decisions. *Transportation Research Part B*, 20(5), 1986, pp. 403–414.
9. Kita, H. A Merging-Giveway Interaction Model of Cars in a Merging Section: A Game Theoretic Analysis. *Transportation Research Part A*, 33(3), 1999, pp. 305–312.
10. Salvucci, D. D., A. and Liu. The Time Course of a Lane Change: Driver Control and Eye-Movement Behavior. *Transportation Research Part F*, 5(2), 2002, pp. 123–132.
11. Webster, N. A., T. Suzuki, M. Kuwahara. Driver Model for Traffic Simulation with Tactical Lane Changing Behavior. *Seisan-Kenkyu*, Vol. 59, No. 3, 2007, pp. 42–45.
12. Toledo, T., H. N. Koutsopoulos, M. Ben-Akiva. Estimation of an Integrated Driving Behavior Model. *Transportation Research Part C*, 17(4), 2009, pp. 365–380.
13. Shvestsov, V., and D. Helbing. Macroscopic Dynamics of Multilane Traffic. *Physical Review E*, Vol. 59, No. 6, 1999, pp. 6328–6338.
14. Hoogendoorn, S. P., and P. Bovy. Platoon-based Multiclass Modeling of Multilane Traffic Flow. *Networks and Spatial Economics*, Vol. 1, 2001, pp. 137–166.
15. Daganzo, C. F. A Behavioral Theory of Multi-lane Traffic Flow. Part I: Long Homogeneous Freeway Sections. *Transportation Research Part B*, Vol. 36, 2002, pp. 131–158.
16. Daganzo, C. F. A Behavioral Theory of Multi-lane Traffic Flow. Part II. Merges and the Onset of Congestion. *Transportation Research Part B*, Vol. 36 (2), 2002, pp. 159–169.
17. Hong, S., S. Tanaka, and M. Kuwahara. Study on an Effectiveness-Evaluation Method of Dynamic Channelization on Merging Sections of Tokyo Metropolitan Expressway. *Seisan-Kenkyu*, Vol. 62(2), 2010, pp. 158–163 (in Japanese).
18. Okaue, M., and M. Okushima. Real-Time Traffic Simulation Model for Traffic Control with Accident Urban Expressway. *Journal of Japan Society of Civil Engineers D3*, Vol. 67, No. 5, 2011, pp. I_1071-I_1078 (in Japanese).
19. Tang, T. Q., S. C. Wong, H. J. Huang, and P. Zhang. Macroscopic Modeling of Lane-Changing for Two-Lane Traffic Flow. *Journal of Advanced Transportation*, 43(3), 2009, pp. 245–273.
20. Jin, W. L. A Kinematic Wave Theory of Lane-Changing Traffic Flow. *Transportation Research Part B*, Vol. 44, 2010, pp. 1001–1021.
21. Jin, W. L. A Multi-Commodity Lighthill-Whitham-Richards Model of Lane-Changing Traffic Flow. *Transportation Research Part B*, Vol. 57, 2013, pp. 361–377.
22. Sheffi, Y. *Urban Transportation Networks: Equilibrium Analysis with Mathematics Programming Methods*. Prentice Hall, Englewood Cliffs, N.J., 1984.
23. van Lint, J. W. C., S. P. Hoogendoorn, M. Schreuder. FASTLANE: New Multiclass First-Order Traffic Flow Model. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2088, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp.177–187.
24. Wang, Y., and M. Papageorgiou. Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter—A General Approach, *Transportation Research Part B*, Vol. 39, No. 2, 2005, pp. 141–167.
25. Patire, A. D., and M. Cassidy. Lane-Changing Patterns of Bane and Benefit: Observations of an Uphill Expressway. *Transportation Research Part B*, 45(4), 2011, pp. 656–666.

CALIBRATION

Heterogeneous Nonlinear Car-Following Laws for Traffic Oscillation Prediction

CHRISTINE RHOADES

XIN WANG

YANFENG OUYANG

University of Illinois at Urbana–Champaign

INTRODUCTION

Traffic oscillations have been the subject of intensive research since the 1950s. Originally, linear car-following laws were used to analyze oscillation properties (1, 2), but more recently, nonlinear car-following laws have been used in order to reproduce more accurate car-following behavior (3). While calibrating the nonlinear laws with field data, however, most research focuses on matching vehicle trajectories in the time domain. Li et al. (4, 5) incorporated frequency domain properties (e.g., oscillation periodicity and amplitude) into car-following model calibration such that the describing-function approach (DFA) can be used to predict the oscillation behavior of a homogeneous platoon (6). No consideration was given to ensure prediction accuracy in both the time- and frequency-domains simultaneously.

This paper proposes a new approach to calibrate the parameters in a driver-dependent nonlinear car-following law based on field trajectories. This method is shown to achieve a better balance between time- and frequency-domain trajectory properties. The developed calibration framework implements maximum likelihood estimation with a simulation-based feedback incorporating both time- and frequency-domain prediction errors. The likelihood estimator is obtained from a modified Tobit model to capture the nonlinearity of the car following model (e.g., because of truncation near zero or maximum speeds). The feedback is established by comparing the observed field trajectory with the simulated one under a certain car-following law, where their actual trajectories and frequency spectrums are used to compute their time- and frequency- domain errors, respectively. The car-following law is calibrated for each pair of leader-follower drivers, allowing researchers to consider a platoon of heterogeneous drivers. The DFA can then be used to predict oscillation propagation in a platoon.

METHODOLOGY

Researchers consider n vehicles, numbered $i = 1, 2, \dots, n$, traveling in a platoon in one lane. The trajectory of vehicle i , with position at time t denoted by $x_i(t)$, is assumed to satisfy a nonlinear car-following law. Field trajectory data is used to calibrate the car-following law, and the designed feedback guarantees that the simulated trajectories reproduce the field trajectory well in the time–space diagram and the frequency spectrum. The DFA can then be used to predict the oscillation propagation for the entire platoon, which serves as a validation of the calibration approach. In the rest of this section, the paper starts by describing the model calibration technique used to obtain simulated trajectories from field data.

Model Calibration

For illustration, researchers calibrated Newell's nonlinear car-following law, which adjusts the desired velocity based on spacing and physical bounds (7):

$$v_i^*(t) = \min\{0, k(x_{i-1}(t-r) - x_i(t-r)) - \omega, v_{\max}\}$$

where ω is the backward shockwave speed, k is the sensitivity factor of driving aggressiveness, v_{\max} is the upper bound on velocity, and τ is the reaction time. A modified version of the Type I standard Tobit model (8) is used to calibrate the vector of model parameters $\langle k, \tau, \omega, v_{\max} \rangle$. Since velocity has both lower and upper bounds, the Tobit model must be censored from above and below, yielding the following piecewise model for the actual velocity of vehicle i at time t :

$$0 \quad \text{if } v_i^*(t) \leq 0$$

$$v_i(t) = v_i^*(t) + \varepsilon \quad \text{if } 0 < v_i^*(t) < v_{\max}$$

$$v_{\max} + \varepsilon \quad \text{if } v_i^*(t) \geq v_{\max}$$

where the error term, ε , follows a normal distribution with mean zero and standard deviation σ .

The model is then calibrated using maximum likelihood estimation based on observed trajectories in time increments. Note that the velocity upper bound is one of the maximum likelihood estimators (MLE). To obtain a balance between the time- and frequency-domain properties of the trajectory, two adjustments are made to the MLE calculation. The first adjustment ensures accuracy in the simulated trajectory by introducing a penalty based on the mean-square error between the field and simulated trajectories. The second adjustment ensures accuracy in the frequency domain by introducing another penalty based on the difference in amplitude of the two largest frequencies after the discrete Fourier transform is conducted on both the field and simulated trajectories.

After the adjustments are made, a heuristic search method (e.g., simulated annealing) is used to find the best parameter vector $\langle k, \tau, \omega, v_{\max} \rangle$ that maximizes the likelihood function. Since the platoon is made of heterogeneous drivers, the model calibration process is conducted for each pair of consecutive drivers.

Oscillation Prediction

Given the oscillation properties (i.e., amplitude and periodicity) of the leading vehicle, the DFA (6) can be applied to predict those of the following vehicle under the calibrated nonlinear car-following law. A general car-following law governing any two adjacent trajectories is considered:

$$x_i(t) = L_i[N_i(x_{i-1}(t) - x_i(t))]$$

where $L_i(\cdot)$ is a linear operator and $N_i(\cdot)$ is a nonlinear function of spacing. The trajectory of each vehicle is first decomposed into its macroscopic, \bar{x}_i , and oscillatory, x_i , components. The Fourier

transform is then performed on the oscillatory component, and the describing function method is used to separate the nonlinear and linear components of x_i , yielding

$$x_i(\Omega) \cong L_i(\Omega) \cdot N_i(x_{i-1}(\Omega) - x_i(\Omega)) \cdot (x_{i-1}(\Omega) - x_i(\Omega))$$

in the frequency domain. **Figure 1** shows the system block diagram that represents the above equation, where the linear operator contains an integrator and a time delay (i.e., reaction time).

Li et al. (5) proposed a method to find $x_i(\Omega)$ for a given $x_{i-1}(\Omega)$ in a homogeneous platoon. In this paper, however, the car-following law is calibrated for heterogeneous drivers instead of an average driver, so this process must be repeated for each pair of consecutive drivers. The oscillation properties of the trajectories can then be predicted [see Li et al. (5) for more details] and compared with field trajectories. This confirms the accuracy of the proposed model calibration approach in the frequency domain.

By maintaining a balance between time- and frequency-domain properties, the calibrated car-following models can more accurately reproduce car-following patterns in both domains. The magnitudes of the penalties assigned during the maximum likelihood estimation are adjustable to fit for various scenarios where either the time- or frequency-domain accuracy is more relevant (higher penalty on frequency domain errors would lead to better oscillation propagation predictions).

NUMERICAL EXAMPLE

Empirical trajectory data are used to validate the proposed model calibration and oscillation prediction framework. For illustration, this paper considers the platoon consisting of vehicles 86–101 from the NGSIM data on southbound US-101 in Los Angeles, California, from 7:50 to 8:35 a.m. on June 15, 2005.

Newell's car-following law is used for illustration, and the best Newell model parameters for each driver are calibrated. An example of a calibration result (the 92nd driver) can be seen in **Figure 2**. The calibrated parameters for this driver are $\langle k = 0.7031, \tau = 1.0500 \text{ s}, \omega = 9.5386 \text{ ft/s}, \text{ and } v_{\max} = 47.5789 \text{ ft/s} \rangle$.

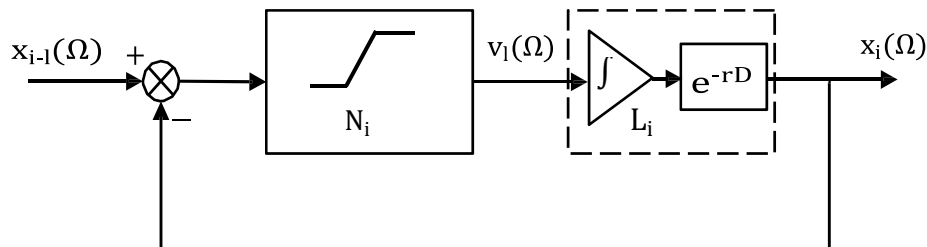


FIGURE 1 Block diagram of Newell's nonlinear car-following law.

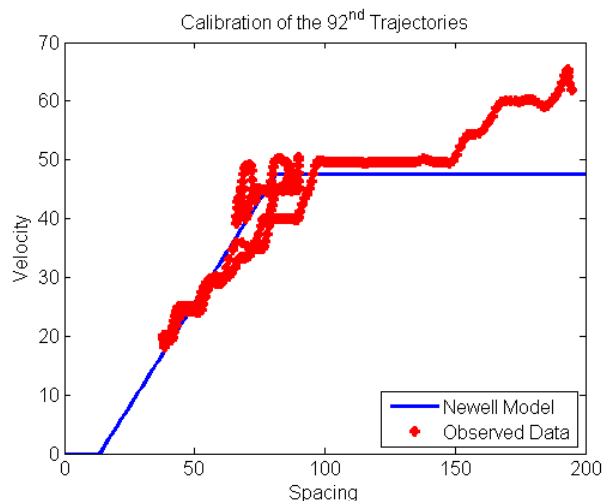


FIGURE 2 Example of a Tobit estimation result (Driver 92).

With the calibrated car-following model for each driver, trajectories of all vehicles in the platoon can be simulated. The simulated and field trajectories match pretty well in the time domain, as shown in Figure 3. Then, the DFA is performed on the simulated trajectories to obtain the oscillation properties; see Figure 4. It can be seen that the predicted growth of oscillation amplitude matches also quite well with field observation in the frequency domain. More examples and insights will be furnished in the full paper.

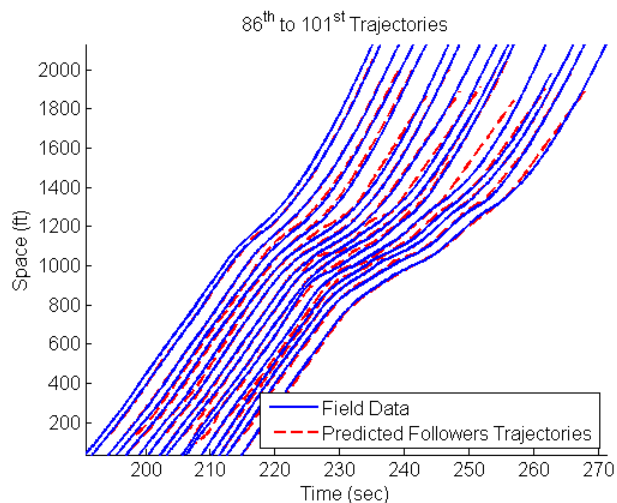


FIGURE 3 Platoon trajectory reproduction prediction in the time domain.

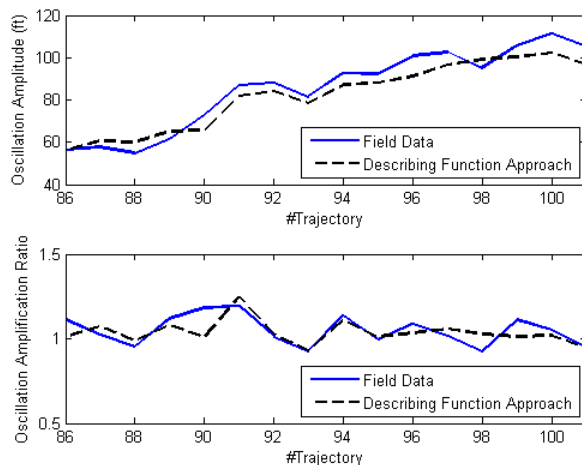


FIGURE 4 Oscillation propagation via frequency–domain DFA.

REFERENCES

1. Chandler, R. E., R. Herman, and E. W. Montroll. Traffic Dynamics: Studies in Car Following. *Operations Research*, 6(2), 1958, pp. 165–184.
2. Herman, R., E. W. Montroll, R. B. Potts, and R. W. Rothery. Traffic Dynamics: Analysis of Stability in Car Following. *Operations Research*, 7(1), 1958, pp. 86–106.
3. Gipps, P. G. 1981. A Behavioural Car-Following Model for Computer Simulation. *Transportation Research Part B*, Vol. 15, 1981, pp. 105–111.
4. Li, X., F. Peng, and Y. Ouyang. Measurement and Estimation of Traffic Oscillation Properties. *Transportation Research Part B*, 44(1), 2010, pp.1–14.
5. Li, X., X. Wang, and Y. Ouyang. Prediction and Field Validation of Traffic Oscillation Propagation under Nonlinear Car-Following Laws. *Transportation Research Part B*, Vol. 46, 2012, pp. 409–423.
6. Li, X., and Y. Ouyang. Characterization of Traffic Oscillation Propagation under Nonlinear Car-Following Laws. *Transportation Research Part B*, 45(9), 2011, pp. 1346–1361.
7. Newell, G. F. Nonlinear Effects in the Dynamics of Car Following. *Operations Research*, 9(2), 1961, pp. 209–229.
8. Amemiya, T. Tobit Models: A Survey. *Journal of Econometrics*, 24(1-2), 1984, pp. 3–61.

CALIBRATION

Using Big Data and Efficient Methods to Capture Stochasticity for Calibration of Macroscopic Traffic Simulation Models

SANDEEP MUDIGONDA

Rutgers, The State University of New Jersey

KAAN OZBAY

New York University

The predictions of a well-calibrated traffic simulation model are much more valid if made for various conditions. Variation in traffic can arise as a result of many factors such as time of day, weather, and accidents. Calibration of traffic simulation models for traffic conditions requires larger datasets to capture the stochasticity in traffic conditions. “Big Data” collected using radio frequency identification (RFID) sensors, smart cellphones, and GPS devices provide extensive traffic data. This study shows the utility of Big Data to incorporate variability in traffic flow and speed for various time periods. However, Big Data poses a challenge in terms of computational effort. With the increase in number of stochastic factors, the numerical methods suffer from the curse of dimensionality. This paper proposes a novel methodology to address the computational complexity due to the need for the calibration of simulation models under highly stochastic traffic conditions. This methodology is based on sparse grid stochastic collocation, which treats each stochastic factor as a different dimension and uses a limited number of points where simulation and calibration are performed. A computationally efficient interpolant is constructed to generate the full distribution of the simulated flow output. This paper shows that this methodology is much more efficient than the traditional Monte Carlo (MC)-type sampling.

INTRODUCTION

Simulation models are mathematical models in which output is derived from a particular model given the input. The input consists of two main groups of data: physical input data, I_s , (e.g., volume counts, capacity, and physical features of roadway sections) and calibration parameters, C_s (i.e., adjustable components of driver behavior). Outputs from a simulation model can be expressed as

$$O_{obs}: f(I_s, C_s) \rightarrow O_{sim}|I_s, C_s + \varepsilon \quad (1)$$

where

- $f(I_s, C_s)$ = functional specification of the internal models in a simulation system;
- O_{sim} = simulation output data given the input data and calibrations;
- ε = margin of error between simulation output and observed field data; and
- O_{obs} = observed field data.

The process of calibration entails adjusting the calibration parameters (C_s) so that the error between the output from simulation and field conditions is acceptable.

Calibration parameters estimated using limited samples are not always representative of all possible conditions of the simulated system and will thus result in inaccurate predictions. In other words, models that are not adequately calibrated cannot accurately capture time-varying conditions of traffic. Traditional sources of traffic data used in the calibration of traffic models are either limited by the availability of the data that only cover typical conditions or may not be reliable enough. However, with the advent of new technologies, information is on the fingertips of users by means of smart phones, GPS-equipped devices, and RFID readers. The rapid rise in information technology also has resulted in innovative ways to obtain space- and time-sensitive information in real time. This, in turn, has led to massive amounts of passively collected location and event data for various time periods, also called “Big Data.” Big Data provides an opportunity to validate and calibrate traffic simulation models for a variety of conditions.

Variability can be incorporated within inputs (demands), I_s , and calibration parameter set (supply), C_s , during different periods of the day, weather conditions, driver population composition, highway geometry, and more. There were a number of studies that captured traffic variability (1–5). However, the increase in the number of factors affecting stochasticity increases the dimensionality of the calibration process. This in turn results in increased computational effort required in calibrating traffic simulation models for different conditions such as variability within weekday/weekend, and seasonal variability, and special situations including adverse weather and workzones.

This study proposes a novel methodology to address the computational complexity due to the need for the calibration of simulation models under highly stochastic traffic conditions. The utility of Big Data to incorporate variability in traffic flow and speed for various time periods is shown.

LITERATURE AND MOTIVATION

There are myriad studies that deal with calibration of traffic simulation models (5–11, 12–19). Because of space constraint, only a sample of them is shown in [Table 1](#).

Table 1 also shows the data used in these studies for the calibrating process. It can be seen that in most studies data used for calibration are limited to a.m. and p.m. peak periods no more than a few days. Thus, the data capture only a few specific conditions or are a dilute sample of different conditions. Hence, it is expected that the model predictions will only be accurate for those specific conditions.

The effect of data and parameter uncertainty in traffic simulation models has received considerable attention recently (18, 19). Studies from other fields indicate that bias and variance in simulation output results are a result of the bias and variance in the input models used, after simulation error is eliminated; the input models consist of simulation model inputs and parameters.

TABLE 1 Summary of Literature on the Calibration of Traffic Simulation Models

Authors	Complexity; Simulation Tool	Type of Roadway Section	Performance Outputs	Data Used in Calibration
Ma and Abdulhai (7)	Micro; PARAMICS	Urban	Traffic counts	Detector data for 1 h during a.m. peak
Kim and Rilett (6)	CORSIM, TRANSIMS	Freeway	Volume	Data 5 loop detector stations for 13.9-m section of freeway for 1 h during a.m., p.m., and off peak
Hourdakis et al. (11)	Micro; AIMSUN	Freeway	Volume	5-min data from 21 detector stations for a 12-m freeway section during p.m. peak for 3 days
Jha et al. (12)	Micro; MITSIMLab	Urban network	Travel time	Detector data for 15 days for a.m. and p.m. peaks on a large urban network
Toledo et al. (13)	Micro; MITSIMLab	Freeway, arterial	Speed, density	Data from 68 detector stations on 3 freeways for 5 weekdays
Qin and Mahmassani (14)	Macro; DYNASMA RT-X	Freeway network	Speed	Data from 7 detector stations on 3 freeways during a.m. peak for 5 weekdays
Balakrishna et al. (15)	Micro; MITSIMLab	Freeway, parkway	Traffic Counts	15-min data from 33 detector stations
Zhang et al. (8)	Micro; PARAMICS	Urban freeway network	Flow, Occupancy	5-min detector count during p.m. peak for 7 days
Li et al. (1)	Macro	Freeway	Flow	Loop detector data
Lee and Ozbay (5)	Micro; PARAMICS	Freeway	Speed, Counts	5-min detector count during a.m. peak for 16 days
Zhong and Sumalee (3)	Macro	Freeway	Flow	Loop detector data for 7 h on 3 days in 2 years
Yang and Ozbay (9)	Micro; PARAMICS	Freeway	Traffic conflict, lane change, volume, and speed	NGSIM trajectory data for US-101 for 15 min
Henclewood et al. (18)	Micro	Freeway	Travel time distribution	NGSIM trajectory data for Peachtree Street in Atlanta, Ga., for 30 min
Punzo et al. (19)	Micro	Freeway	Speed	NGSIM trajectory data for I-180

FHWA's *Traffic Analysis Toolbox* (20) recommends that if $GEH < 4$, where

$$GEH = \sqrt{\frac{\frac{1}{T} \sum_{i=1}^T (O_{sim,i} - O_{obs,i})^2}{\frac{1}{2T} \sum_{i=1}^T (O_{sim,i} + O_{obs,i})}}$$

for link volumes for 85% of the links and average travel times are within 15% of observed values, then it is considered as a satisfactorily calibrated model (20). In order to achieve this level of calibration for various conditions (peak, off-peak, weekends, normal and inclement weather, under accident, and other events), a detailed level of data is required.

As illustrated in Table 1, most of the studies in traffic simulation used a limited amount of data focusing on a small set of conditions or time periods, or both. As depicted in Figure 1, using only smaller samples of data will not accurately capture variation in traffic data [additionally, the sources of field data in most of these studies, except (5–9, 19), have been traditional sources such as loop detectors or manually-acquired data from captured videos, which can be cumbersome and not always accurate enough]. Using these models for conditions other than the ones for which calibration data are available would not yield accurate results.

Ozbay et al. (22) showed that the existence of a typical day in traffic demand is not always likely. Hence, to obtain accurate predictions from a traffic simulation model, it is important to consider not only the demand from various clusters, but also the variation of demand within each cluster.

COMPUTATIONAL COMPLEXITY

In cases in which large sources of data spanning different conditions are available, to capture the stochasticity in traffic conditions, there is an increase in number of factors of stochasticity. This in turn increases the dimensionality of the calibration process. This then results in increased computational effort required in calibrating traffic simulation models for various conditions.

Most studies capturing stochasticity in computational traffic models use an MC-type independent sampling of M simulation runs for various traffic conditions. The number of

replications needed to be at a level of precision γ is given by $\left(\frac{t_{M-1, 1-\alpha/2} * S}{\gamma}\right)^2$. However, the

convergence rate for MC-type method or Latin hypercube sampling is slow, $O(1/\sqrt{M})$ (23). All of these numerical methods suffer from the curse of dimensionality. Thus, depending on the size of the network and number of stochastic dimensions, these approaches can become prohibitive in terms of computational effort. It may not be possible at all to simulate the output for each and every possible realization of parameter and input. Also, all possible points in the stochastic space of simulation output may not have the corresponding observed data. Thus, it is important to obtain an effective interpolation methodology for predicting output for points with no data.

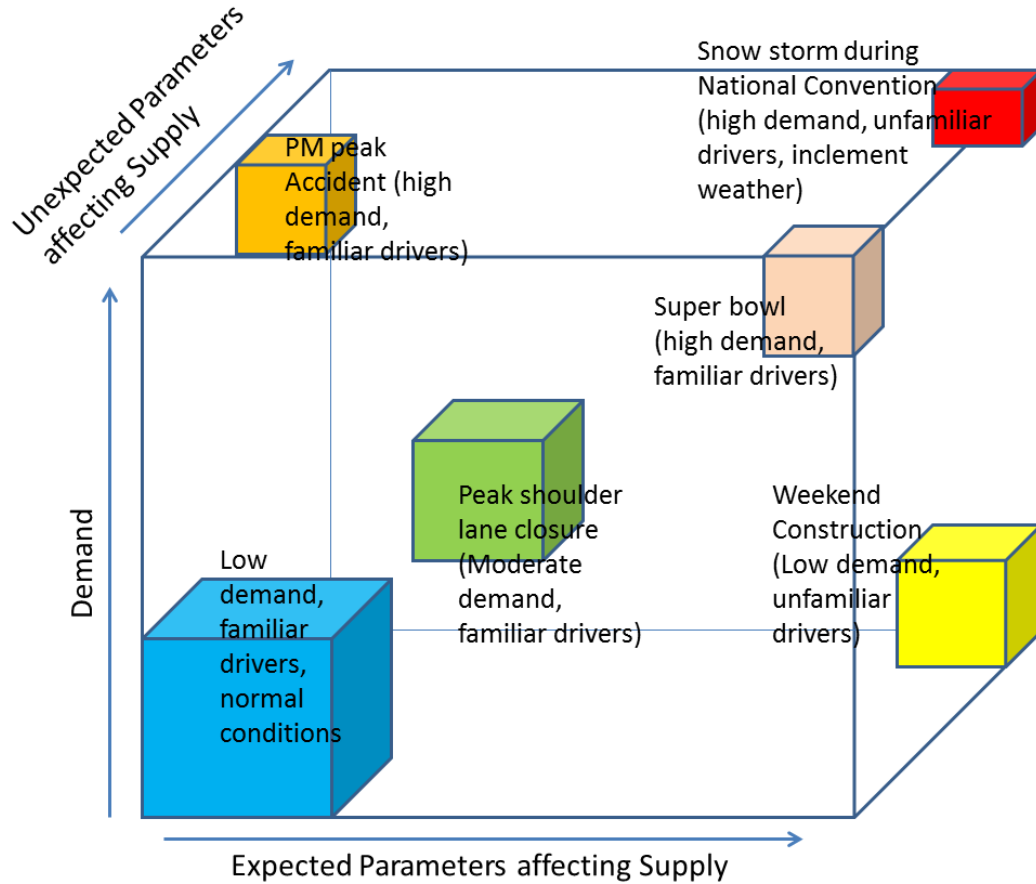


FIGURE 1 Illustration of various traffic conditions for which data are required for calibration [adapted from (21)].

METHODOLOGY

Use of Big Data

Based on the discussions above, making accurate predictions using traffic simulation models requires calibration data from many sources and in great detail. This data need can be effectively addressed by the advent of new technologies such as GPS, cellular phones, and RFIDs. The ubiquity of these technologies ensures that data of great detail and variety are available. Various demand, speed, flow, and event data can be obtained from the Big Data sources.

This study uses a hybrid of electronic toll collection (ETC) data for demand and traffic sensor data for speed and flow. ETC data are collected for all toll ways in the United States and in New Jersey. Taking toll facilities in New Jersey as an example, New Jersey Turnpike (NJTPK) is spread over 150 mi with 28 interchanges and 366 toll lanes. The Garden State Parkway is about 170 mi long with 50 toll plazas and 236 toll lanes. Each freeway carries up to 400,000 vehicles per day (24). The ETC data are collected at toll plazas on these freeways (24). The ETC dataset consists of the individual vehicle-by-vehicle entry and exit time data. It also consists of the information regarding the lane through which each vehicle was processed (both E-ZPass and cash users), vehicle types, number of axles, and more.

The demand from the ETC data is divided into clusters so that different demand patterns can be analyzed separately. The corresponding sensor data are also obtained into clusters. The simulation is performed using the clustered demand data distribution and simulation output of flow, and density is compared to the observed distribution from sensor data.

Stochastic Collocation

As mentioned earlier, there is a need to combat the issue of a high number of replications and dimensionality when using MC-type methods in capturing stochasticity using computational traffic models. Stochastic spectral methods provide an efficient alternative to MC-type methods. In these methods, stochasticity is treated as another dimension, and the stochastic solution space Ω is discretized and approximated (Γ). One such stochastic spectral method is stochastic collocation (SC), where the approximation is performed using deterministic solutions at a set of prescribed nodes (collocation points) and an interpolation function. For a set of Q points in the stochastic space where the simulation output is calibrated, $\{\Theta_j\}_{j=1}^Q \in \Gamma$, the polynomial interpolation can be shown as (25)

$$\begin{aligned} \rho(x, t, \xi) &\approx \int_{\Gamma} \rho(x, t, \xi) p(\xi) d\xi \approx \hat{\rho}(x, \Theta) = \sum_{j=1}^Q \tilde{\rho}_j(x) \Phi_k(\Theta_j), \forall x \in D \\ \Phi_k(\Theta_j) &= \text{interpolating basis polynomials} \\ \tilde{\rho}_j(x) &= \text{the deterministic solution at } \Theta_j \end{aligned} \quad (2)$$

The choice of weights (i.e., the basis functions Φ_k) in SC depends on the interpolation technique. It is important to note that, unlike MC-type sampling methods, the weights assigned to each deterministic output are different in stochastic collocation methods.

Computationally efficient schemes such as the Smolyak algorithm are used to reduce the number of collocation points (26, 27) at higher dimensions of stochasticity. Smolyak algorithm, developed originally for multidimensional integration, entails evaluating deterministic solutions at the nodes of sparse sampling grids and building the interpolation function. One dimensional interpolant, with the Smolyak algorithm, is given by, $U^i(f) = \sum_{j=1}^{m_i} f(\Theta_j) L_j$, $m_i = \#$ nodes at level i .

Sparse interpolant in N -dimensions and q - N interpolation order is given by

$$\begin{aligned} \Delta^i &= U^i - U^{i-1}; U^0 = 0 \\ A_{q,N}(f) &= \sum_{|i| \leq q} (\Delta^i \otimes \dots \otimes \Delta^{i_N})(f) = A_{q-1,N}(f) + \sum_{|i|=q} (\Delta^i \otimes \dots \otimes \Delta^{i_N})(f) \end{aligned} \quad (3)$$

The advantage of this recursive/nested structure is that to increase the order of interpolation (accuracy) researchers can use all of the deterministic solutions from the previous steps, $A_{q-1,N}$, by adding a few more deterministic solutions: Δ_{iN} . When new data are available, additional deterministic solutions can be evaluated, and accuracy of interpolant is improved.

Convergence rate of the interpolant is of the order, $O(Q^{-2} |\log Q|^{3(N-1)})$ (for piecewise linear basis), $O(Q^{-k} |\log Q|^{(k+2)(N-1)})$ (for k -polynomial basis). This rate can be controlled by the

polynomial order k (26, 27). Thus, it is shown, empirically, that convergence of this interpolant is better than the MC method.

An illustration of the discretization of the stochastic space is shown in Figure 2.

From each realization of the parameter set, using the demand distribution as an input, the simulation output distribution (e.g., flow or density distribution) is generated. This distribution is compared with the observed output distribution and using a test statistic (such as the test statistic from the KS test), the error is estimated. This error is used as an objective function and is minimized as part of the multi-objective parameter optimization (shown in Equation 4) using the simultaneous perturbation stochastic approximation (SPSA) algorithm. In this study the weights (w) used are the coefficients of variation of each output measure from the observed data.

$$\min_{\Theta_t} \sum_{i=1}^N \{w_1 U_1(q_i^{Ob}, q_i^S(\Theta_t^k)) + w_2 U_2(\rho_i^{Ob}, \rho_i^S(\Theta_t^k))\} \quad (4)$$

where

- q_i^{Ob}, q_i^S = observed and simulated flows at location i ;
- ρ_i^{Ob}, ρ_i^S = observed and simulated densities for location i ;
- Θ_t^k = parameter set for time period t and iteration k ;
- w_1, w_2 = weights for the error measures; and
- U_1, U_2 = functions representing the errors in flow and density.

The main advantages of using this methodology are the following:

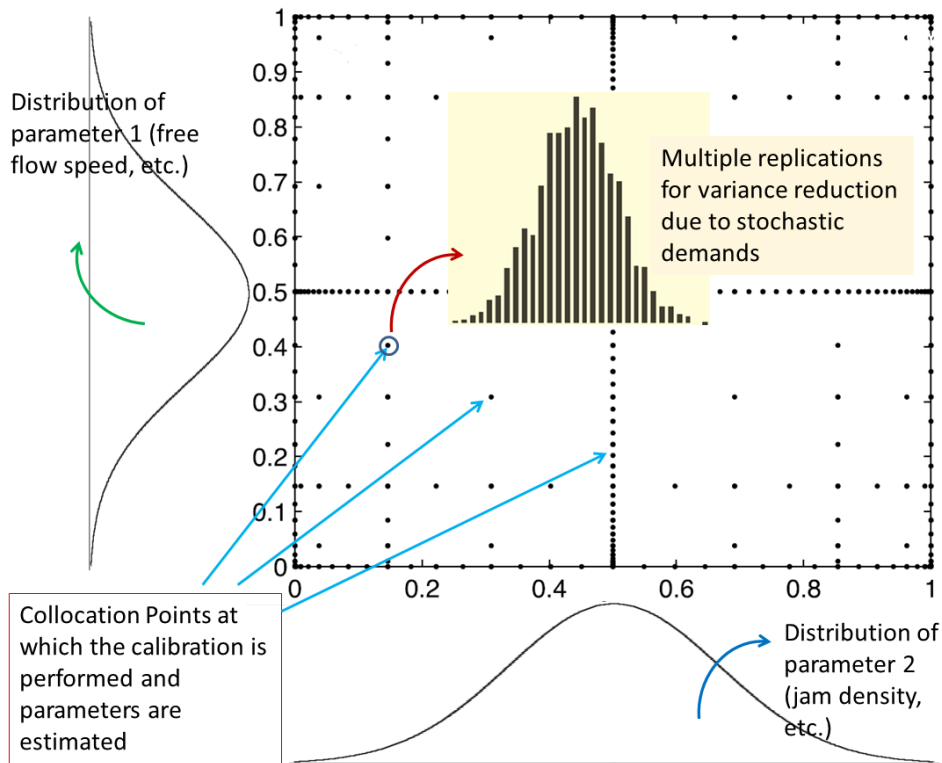


FIGURE 2 Example of approximation of stochastic space by collocation points.

1. Flexibility in applying to any type of traffic simulation (first order, second order, meso–microscopic, and more);
2. Computationally more efficient than MC-type exhaustive sampling methods with effective interpolant to generate full distribution of simulation output;
3. Time consumed by the collocation approach can be further reduced by parallelizing the simulation under each condition; and
4. Nested form of the algorithm is useful in refining the interpolant as and when there is new data available.

RESULTS

In order to illustrate the stochastic variation in traffic conditions, a three-lane section of the NJTPK at Interchange 7 was chosen. Although, microscopic traffic simulation tools such as Paramics provide a detailed and relatively accurate platform for modeling, the model building, calibration, and execution can be very time consuming. When studying the effects of various stochasticities, the focus was on a first-order macroscopic traffic simulation model to model the traffic flow in the section. The time and space for the model were discretized using the cell transmission model. A schematic representation of the simulated section, the stochastic input, and model parameters are shown in Figure 3.

The variation in demand at this section was captured using the ETC data for every 5 min between January 1 and August 31, 2011. The demand is divided into clusters using *k*-means algorithm. The optimal clusters for the demand distribution are shown in Figure 4. These data illustrate that primarily there are three clusters of data: weekday, weekend high, and weekend low.

Each line depicted in Figure 4 is the mean of the cluster for each of the demand cluster. Using the actual demand values from each of the mean does not take into account the variability in traffic conditions and may not be useful. Thus, for each cluster, the distribution of demand during each 5-min time period is generated as illustrated in Figure 5. The simulation is performed for a.m. peak (7:00 to 9:00 a.m.) and off-peak (10:00 a.m. to 12:00 p.m) period for weekday demand cluster and peak period (10:00 a.m. to 12:00 p.m.) of the weekend demand

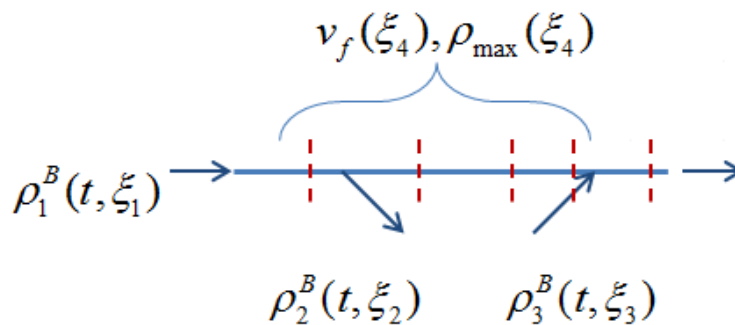


FIGURE 3 Schematic representation of the study section.

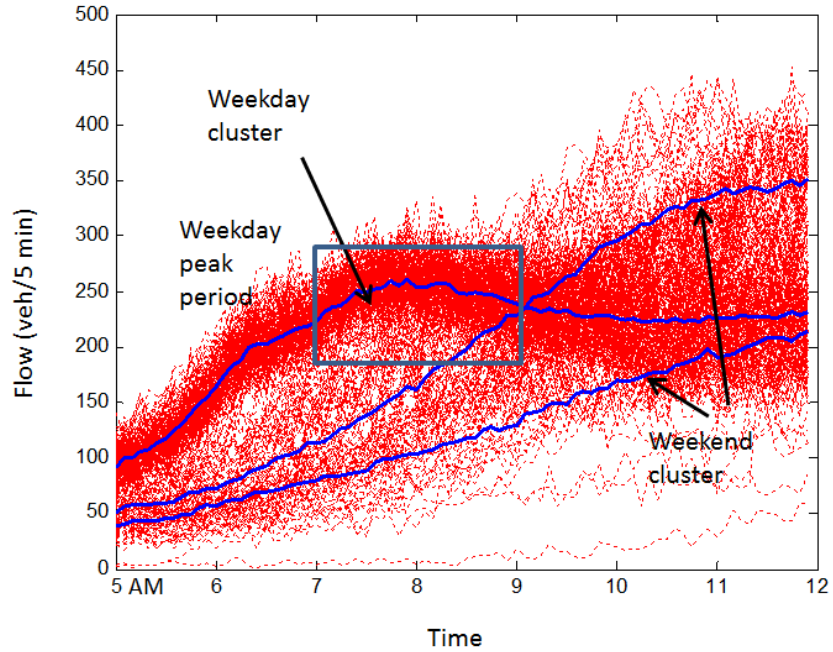


FIGURE 4 Clusters and distribution of demand for the mainline section at Interchange 7 of NJTPK.

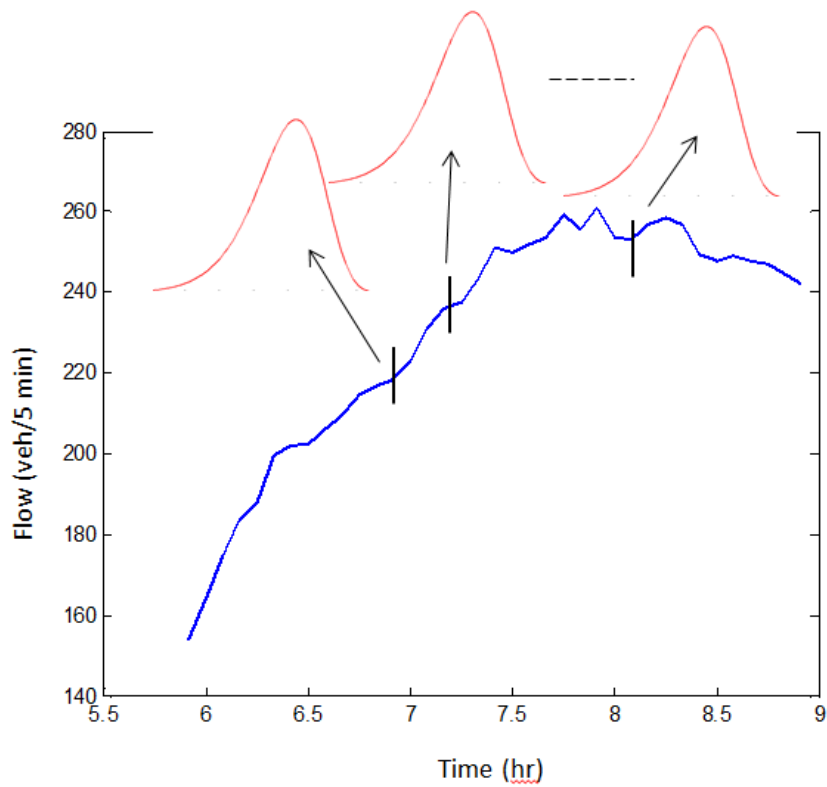


FIGURE 5 Illustration of demand distribution at various sampled times during a.m. peak period.

cluster. It is to be noted that, especially during the peak period, the demand distribution for each sampling period (15 min) is different from the other period. Additionally, since this input parameter is demand as opposed to flow, each sampling period can be treated as independent of the other time period. So for simulating 7:00 to 9:00 a.m. with 15-min demands, there will be eight separate demand distributions, each of which has to be treated as an independent stochastic dimension.

Traffic sensor data available between January 1, 2011, and August 31, 2011, for every 5 min were used to generate the speed and flow data required for calibration. The sensor data were separated into the same clusters for the same time periods as the demand clusters. Thus, similar flow distribution data were generated.

For each of the three time periods, a.m. peak (7:00 to 9:00 a.m.) and off-peak (10:00 a.m. to 12:00 p.m.) periods for weekday demand cluster and peak period (10:00 a.m. to 12:00 p.m.) of the weekend demand cluster, the simulation model was calibrated. As mentioned in the methodology, with the demand distribution as an input, for each realization of the parameter set, the simulation output flow distribution was generated. This distribution was compared with the observed flow distribution, and using a test statistic [such as the test statistic from the Kolmogorov-Smirnov test (KS) test], the error was estimated. This error was used as an objective function and was minimized using the SPSA algorithm. The result of calibration was demonstrated using the comparison of simulated and observed flow.

For this study, the Clemshaw-Curtis grid (a two-dimensional version of which can be seen in Figure 2) was the appropriate sparse grid to discretize the stochastic demand. The simulation was calibrated using the demand values at each of these grid points. The objective function for calibration was the test statistic used in the KS test at 90% significance, maximum separation between two distributions. As mentioned in Equation 3, a sparse grid interpolation was performed for the output of the simulation and a Smolyak interpolant is constructed. Distribution of simulated flows was obtained by repeated evaluation of the Smolyak interpolation function. The simulated flow distribution was compared to the observed distribution from the sensor data.

The comparison of observed and simulated flow distributions from the calibrated model for a.m. peak period is shown in Figure 6.

In order to compare the efficiency of the stochastic collocation approach, the distribution of simulated flow after model calibration was also generated using the MC sampling method. In order to achieve the flow distribution, the SC approach required 2,433 evaluations for various stochastic demand combinations. However, using an MC-type sampling required 240,000 runs of the simulation model. The reason, as mentioned earlier, was because of the ability to construct an efficient Smolyak interpolant that uses the simulation output from much fewer runs.

As the second case study, the flow distribution during the off-peak period was also generated using the sparse grid interpolation. The comparison of observed and simulated flow distribution from the calibrated model for weekday off-peak period is shown in Figure 7. It was observed that there was no flow breakdown during this time period. As in the first case study, the distribution of simulated flow after model calibration was also generated using the MC sampling method. In order to achieve the flow distribution, the SC approach required 441 evaluations for various stochastic demand combinations. However, when using an MC-type simulation required 5,420 samples.

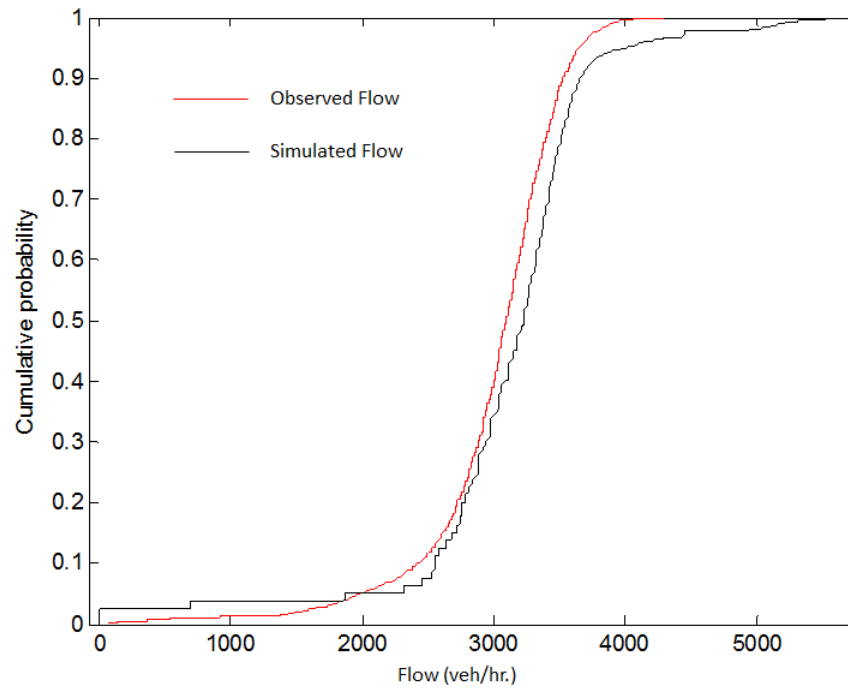


FIGURE 6 Comparison of observed and simulated link flow distributions during a.m. peak period.

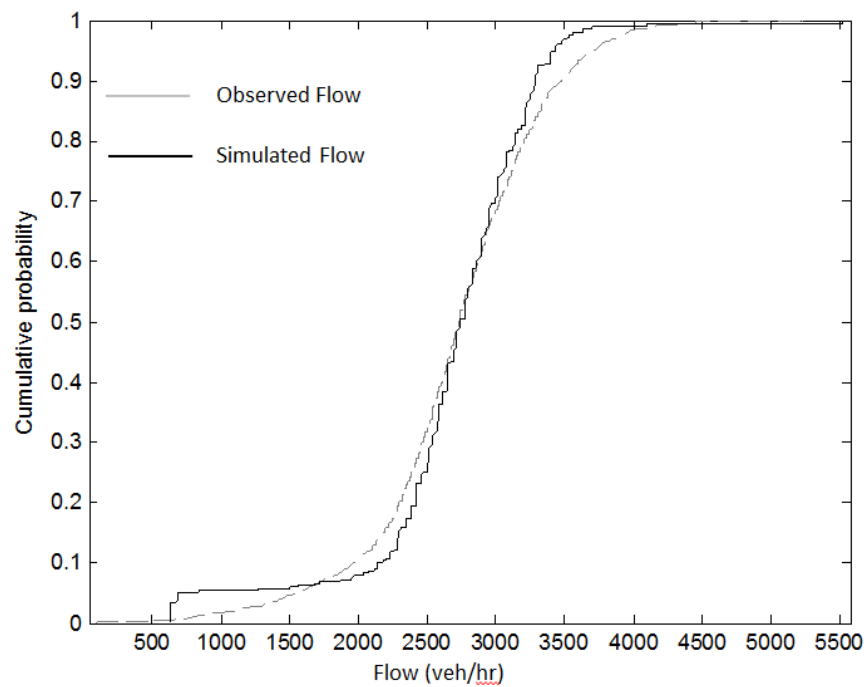


FIGURE 7 Comparison of observed and simulated link flow distributions during off-peak period.

As the third case study, the flow distribution during the high- and low-demand weekend peak period was also generated using the sparse grid interpolation. The comparison of observed and simulated flow distribution for weekend peak period is shown in Figure 8. It can be observed that there was no flow breakdown during this time period. Distribution of simulated flow is also generated using the MC sampling method. In order to achieve the flow distribution, the SC approach required 441 evaluations for various stochastic demand combinations. However, using an MC-type simulation required 8,000 samples.

The motivation behind using data from a variety of conditions was to capture the stochasticity in traffic conditions. To illustrate the drawback of using limited data, researchers compared the distribution of flow for high- and low-weekend demands with the case where only three weekend days of flow and demand were used to calibrate the weekend model. The simulated flow distributions (shown in Figure 8) from a limited data model do not match, not only the weekend model with high demand but also the weekend model with low demand. This illustrates the drawback in using limited data for model calibration and the importance of considering stochasticity in traffic conditions when calibrating traffic simulation models.

CONCLUSIONS AND FUTURE WORK

The predictions of well-calibrated traffic simulation model are robust if the predictions made for various conditions are accurate. Variations in traffic can arise as a result of many factors, such as time of day, weather, and accidents. Calibration of simulation models for traffic conditions requires larger than traditionally adopted datasets capturing the stochasticity in traffic conditions.

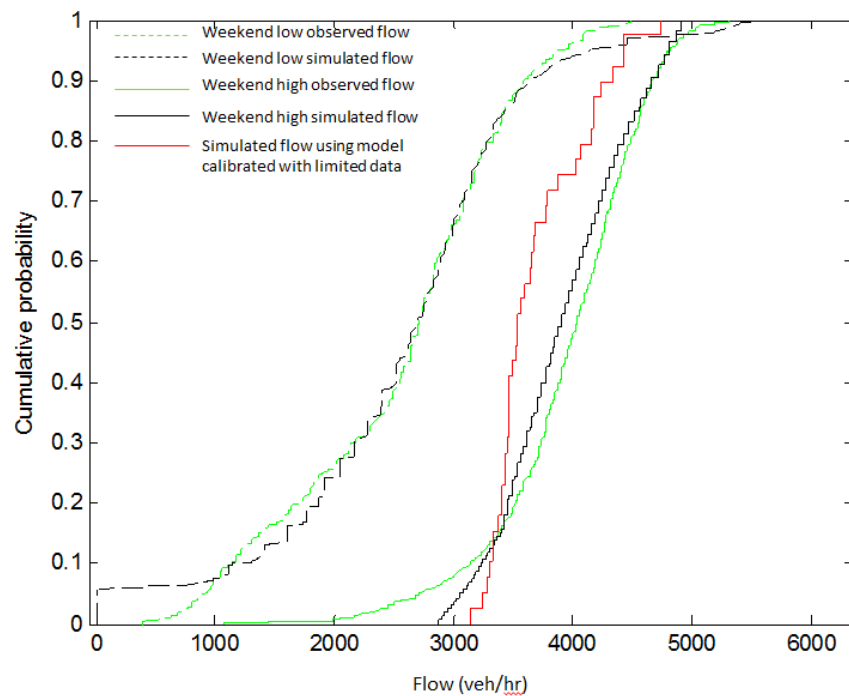


FIGURE 8 Comparison of observed and simulated link flow distributions for low and high clusters during weekend peak period.

The advent of new technologies such as RFID sensors, smart cellphones, and GPS devices provides extensive traffic data, an example of Big Data. Big Data provides a viable alternative to traditional traffic data sources in providing larger and richer datasets. Although Big Data provides greater variation in data, it poses a challenge in terms of computational effort. With the increase in number of stochastic factors, the numerical methods suffer from the curse of dimensionality. If traditional MC-type sampling is used, the computational effort required to simulate and calibrate traffic simulation models for various conditions could become intractable.

This study used ETC data for a period from January to August 2011 to capture various traffic conditions. A novel methodology to encapsulate stochasticity into macroscopic traffic simulation models and their calibration with much lower computational effort was proposed. Stochastic collocation, a type of stochastic spectral method, was used to capture stochasticity in traffic. This method treats each stochastic factor as a separate dimension. Each dimension is discretized using a set of collocation points, and an interpolant for the output is constructed using the simulation output at these points. In this study researchers used the Smolyak sparse grid interpolation method due to the high number of stochastic dimensions.

The main advantages of using this methodology are the following:

1. Flexibility in applying to any type of traffic simulation (first order, second order, meso–microscopic, and more);
2. Computationally more efficient than MC-type exhaustive sampling methods with effective interpolant;
3. Time consumed by the collocation approach can be further reduced by parallelizing the simulation under each condition; and
4. Nested form of the algorithm is useful in refining the interpolant as and when there are new data available.

To demonstrate the usefulness of this methodology, it was tested for a short on-ramp–off-ramp section of NJTPK in the vicinity of Interchange 7. The variation in demand at this section was captured using the ETC data for every 5 min between January 1, 2011, and August 31, 2011. The demand was divided into clusters using *k*-means algorithm into weekday, weekend high, and weekend low clusters. The methodology was applied to calibrate the macroscopic first order traffic simulation model for a.m. peak (7:00 to 9:00 a.m.) and off peak (10:00 a.m. to 12:00 p.m.) period for weekday demand cluster and the peak period (10:00 a.m. to 12:00 p.m.) for high and low weekend clusters. For calibrating the model, researchers used the test statistic from the KS test for flow distributions on the link as the objective function. This measure was minimized using the SPSA algorithm. It showed that the comparison of simulated and observed flow distributions for the peak and off-peak periods match well. Also compared was the simulated and observed flow distributions for weekend peak demand; they matched well. Also illustrated was the advantage of using varied traffic conditions for model calibration by comparing the simulated flow distributions generated using models calibrated with a large set of data and models calibrated using limited days of demand and flow data.

As a part of future work, the probability density function of flows and densities at various sections will be validated using a holdout dataset. Also this methodology will be extended to larger highway sections for larger freeway sections with stochastic flow–density relationship.

REFERENCES

1. Li, J., Q. Chen, H. Wang, and D. Ni. Analysis of LWR Model with Fundamental Diagram Subject to Uncertainties. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
2. Ngoduy, D. Multiclass First Order Model Using Stochastic Fundamental Diagrams. *Transportmetrica*, Vol. 7, No. 2, March 2011, pp. 111–125.
3. Zhong, R., and A. Sumalee. Stochastic Cell Transmission Model for Traffic Network with Demand and Supply Uncertainties. *Proc. 6th International Conference of Traffic and Transportation Studies Congress*, 2008.
4. Jabbari, S. E., and H. X. Liu. A Stochastic Model of Traffic Flow: Theoretical Foundations. *Transportation Research Part B*, Vol. 46, 2012, pp. 156–174.
5. Lee, J.-B., and K. Ozbay. New Calibration Methodology for Microscopic Traffic Simulation Using Enhanced Simultaneous Perturbation Stochastic Approximation Approach. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2124, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 233–240.
6. Kim, K. O., and L. R. Rilett. Simplex-based Calibration of Traffic Micro-Simulation Models with Intelligent Transportation Systems Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1855, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp.80–89.
7. Ma, T., and B. Abdulhai. Genetic Algorithm-Based Optimization Approach and Generic Tool for Calibrating Traffic Microscopic Simulation Parameters. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1800, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 6–15.
8. Zhang, M., J. Ma, S. P. Singh, and L. Chu. *Developing Calibration Tools for Microscopic Traffic Simulation Final Report Part III: Global Calibration, O-D Estimation, Traffic Signal Enhancements and a Case Study*. California PATH Research Report, UCB-ITS-PRR-2008-8, June 2008.
9. Yang, H., and K. Ozbay. Calibration of Micro-Simulation Models to Account for Safety and Operation Factors for Traffic Conflict Risk Analysis. Presented at 3rd International Conference on Road Safety and Simulation, September 14–16, 2011, Indianapolis, Ind.
10. Menneni, S., C. Sun, and P. Vortisch. Microsimulation Calibration Using Speed-Flow Relationships. Making NGSIM Data Usable for Studies on Traffic Flow Theory. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2088, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 1–9.
11. Hourdakis, J., P. G. Michalopoulos, and J. Kottommannil. A Practical Procedure for Calibrating Microscopic Traffic Simulation Models. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1852, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 130–139.
12. Jha, M., G. Gopalan, A. Garms, B. P. Mahanti, T. Toledo, and M. E. Ben-Akiva. Development and Calibration of a Large-Scale Microscopic Traffic Simulation Model. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1876, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 121–131.
13. Toledo, T., M. E. Ben-Akiva, D. Darda, M. Jha, M., and H. N. Koutsopoulos. Calibration of Microscopic Traffic Simulation Models with Aggregate Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No.1876, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 10–19.
14. Qin, X., and H. S. Mahmassani. Adaptive Calibration of Dynamic Speed-Density Relations for Online Network Traffic Estimation and Prediction Applications. In *Transportation Research Record: Journal of the Transportation Research Board*, No.1876, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 82–89.
15. Balakrishna, R., C. Antoniou, M. Ben-Akiva, H. N. Koutsopoulos, and Y. Wen. Calibration of

- Microscopic Traffic Simulation Models: Methods and Application. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1999*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 198–207.
16. Li, J., Q. Chen, H. Wang, and D. Ni. Analysis of LWR Model with Fundamental Diagram Subject to Uncertainties. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
 17. Sumalee, A., R. X. Zhong, T. L. Pan, and W. Y. Szeto. Stochastic Cell Transmission Model (SCTM): A Stochastic Dynamic Traffic Model for Traffic State Surveillance and Assignment. *Transportation Research Part B*, 2010.
 18. Henclewood, D., W. Suh, M. O. Rodgers, and M. Hunter. Statistical Calibration for Data-Driven Microscopic Simulation Model. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.
 19. Punzo, V., M. Montanino, and B. Ciuffo. Goodness of Fit Function in the Frequency Domain for Robust Calibration of Microscopic Traffic Flow Models. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington D.C., 2013.
 20. Luttrell, T., W. Sampson, D. Ismart, and D. Matherly. *Traffic Analysis Toolbox. Volume VII: Predicting Performance with Traffic Analysis Tools*. Publication FHWA-HOP-08-055. FHWA, U.S. Department of Transportation, 2008.
 21. Wunderlich, K. *Seattle 2020 Case Study, PRUEVIN Methodology*. Mitretek Systems.
 22. Ozbay, K., S. Mudigonda, E. Morgul, H. Yang, and B. Bartin. Big Data and the Calibration and Validation of Traffic Simulation Models. Presented at 93rd Annual Meeting of the Transportation Research Board, Washington, D.C., 2014.
 23. Loh, W.-L. On Latin Hypercube Sampling. *The Annals of Statistics*, Vol. 24, No. 5, 1996, pp. 2058–2080.
 24. New Jersey Turnpike Authority. <http://www.state.nj.us/turnpike/>. Accessed December 25, 2013.
 25. Xiu, D., and J. S. Hesthaven, High Order Collocation Methods for the Differential Equation with Random Inputs. *SIAM Journal of Scientific Computing*, Vol. 27, 2005, pp. 1118–1139.
 26. Ganapathysubramaniam, B., and N. Zabaras. Sparse Grid Collocation Schemes for Stochastic Natural Convection Problems. *Journal of Computational Physics*, Vol. 225, 2007, pp. 652–685.
 27. Klimke, W. A. *Uncertainty Modeling Using Fuzzy Arithmetic and Sparse Grids*. PhD dissertation. Universität Stuttgart, Shaker Verlag, Aachen, 2006.

Empirical Observations of Traffic Analysis

EMPIRICAL OBSERVATIONS OF TRAFFIC ANALYSIS

Continuous Flow Metering Alternative Solution to Alleviate Congestion on Interstate 70 Eisenhower–Johnson Memorial Tunnel

SUSI MARLINA

URS

BRUCE N. JANSON

University of Colorado, Denver

SAEED SOBHI

Colorado Department of Transportation

Metering aims to improve traffic flow on the main lanes of a freeway by regulating the inflow to these lanes. While metering is most often implemented at on-ramps to a freeway, it can also be used to regulate the flow of traffic on the main lanes themselves. Although counterintuitive that regulating flow on the main lanes can also improve overall traffic flow, this can happen in cases where downstream bottlenecks exist. Alternatively, upstream metering may simply be required to prevent excessive queuing at a downstream bottleneck. This paper evaluates two metering techniques applied to the eastbound main lanes of Interstate 70 at the Eisenhower–Johnson Memorial Tunnel (EJMT) to potentially improve traffic flow, shorten the queues both east and west of the EJMT, and reduce the need to completely stop traffic at the EJMT. One strategy evaluated is traditional, or manual, metering in which operations staff stop and start traffic as seen necessary to reduce a downstream queue. The other strategy is called continuous flow metering (CFM), which automatically displays red and green lights at a given cycle length to smooth out the traffic stream as well as reduce downstream queuing. The study results show that CFM did reduce travel times and save person hours of travel in several trial time periods.

INTRODUCTION

Interstate 70 (I-70) provides the only interstate connection from eastern to western Colorado and passes through difficult mountain terrain and physical constraints, including the narrow EJMT, with only two lanes in each direction. I-70 experiences heavy traffic and subsequently high delays, particularly during the winter time, when travelers are returning from the high country ski areas to the Denver metropolitan area. I-70 is classified as an uninterrupted freeway facility. It frequently serves 50,000 vehicles per day and close to 5,000 vehicles during peak hours. I-70 experiences significant regional and national use by commercial and recreational travel. Weekend travelers returning to Denver from the high country often encounter significant traffic jams, delaying trips by hours because of weekend recreational activities in the mountains.

Figure 1 shows the study area: a 16-mi stretch of highway running from Silverthorne at the western end [mile post (MP) 205] to Bakerville at the eastern end (MP 221). There are



FIGURE 1 Study area.

currently three eastbound (EB) lanes from Silverthorne to the lower runaway truck ramp, then two lanes to the US-6 Loveland on-ramp, then three lanes to Hermans Gulch, and then two lanes for the remaining distance to Bakerville. This EB section of highway includes four on-ramps and four off-ramps, none of which lies between Silverthorne and the EB entrance to the EJMT.

The EJMT is the highest highway tunnel in the world (reaching an elevation of 11,000 ft) and is located on I-70 approximately 60 mi west of the Denver metropolitan area. It is a major bottleneck during high-volume periods during winter and summer weekends. Delays are primarily a result of heavy directional volumes, steep grades (7% approaching the EJMT in the EB direction), and a long tunnel (1.7 mi). Slower moving trucks on the up and down grades also contribute to delays, which can be made worse by poor weather conditions and traffic incidents when they occur. In general, a freeway can operate at full capacity when there is good weather, good visibility, no incidents, no work zone activity, and no pavement deterioration serious enough to affect operations.

The basic freeway section analysis of the *Highway Capacity Manual 2010* (HCM) provides little guidance for estimating tunnel lane capacities, since the tunnels affect travel speeds quite dramatically (1). To approximate the capacity of the EJMT, the following specifications were used in the basic freeway analysis of the HCM 2010: 10% heavy vehicles, including trucks, buses, and recreational vehicles; 0 interchanges per mile; 60 mph base free-flow speed; and rolling terrain for two lanes in each direction. However, if the lane and shoulder width assumptions reduce the free-flow speed to below 55 mph, then the HCM does not report a level of service (LOS).

Using these parameters in the HCM basic freeway analysis results in an adjusted capacity of 1,800 passenger cars per hour per lane (pc/h/ln), which is roughly the break point between LOS E and F for constrained lane and shoulder width conditions. Although the HCM does not have a capacity adjustment for tunnel effects, the standard capacity reduction for tighter urban

streets versus open suburban roadways is 10%. Thus, a 10% reduction in capacity to 1,620 pc/h/ln seems about right to account for tunnel effects. This estimate also concurs with the research literature. Levinson et al. (2) estimated the capacity of the Callahan Tunnel in Boston to be between 1,600 and 1,650 pc/h/ln after installing traffic management improvements. Levinson et al. (2) also cite a New York Port Authority estimate of 1,660 pc/h/ln as the maximum theoretical capacity of a tunnel lane. However, observed maximum volumes in New York and New Jersey tunnels suggest a maximum practical capacity of 1,350 to 1,450 pc/h/ln. Lin et al. (3) estimated the capacity of a tunnel in Taiwan after improvements to be 1,300 vehicles per hour per lane (vphpl) in the southbound direction, but only 1,150 vphpl in the northbound direction. Koshi et al. (4) observed the capacities of tunnels in Japan under congested conditions to be in the range of 1,100 to 1,400 pc/h/ln with the average being about 1,325 pc/h/ln.

Although these estimates are wide ranging, based on the above investigations, it was estimated that for normal operations, the capacity of the EJMT is roughly 1,600 vphpl, or 3,200 vph total for the two lanes in each direction.

TRADITIONAL METERING VERSUS CONTINUOUS FLOW METERING

At times, queues of vehicles traveling EB on the east side of the EJMT extend inside the tunnel causing stop-and-go conditions and restricting access inside the tunnel for emergency response vehicles. To alleviate this congestion and public safety concern, the Colorado Department of Transportation (DOT) implemented a CFM strategy versus the traditional metering approach. The CFM approach releases the EB traffic on I-70 just before entering the EJMT at a rate based on freeway volumes and speeds. This approach differed from the traditional metering approach at the EJMT, which stopped all traffic for up to 20 min at a time. The EJMT staff imposed traditional metering at the west entrance to the tunnel to prevent queues in the EB lanes beyond the tunnel from backing up into the tunnel. The shockwave begins in the EB lanes downhill of the tunnel where the geometric characteristics of the highway cause a long queue to form. A study by Zang and Shen (5) found that a lane-drop with low remaining capacity, high travel demand, and a short distance between the merging sign and taper is most likely to produce high oscillations in flow. Furthermore, Laval et al. (6) found that the evolution of oscillations occurred in freeway segments with no entrances or exits. In this study, oscillation occurs with no entrances or exits, but because of a lane drop, high traffic volumes, traffic incidents, and frequently poor weather conditions.

The traditional metering scheme is to completely stop EB traffic at the west entrance and to allow the queue east of the tunnel to dissipate to some point past the US-6/Loveland pass on-ramp (MP 216.5). At that point, the metered traffic is released. The process is repeated if the EB queue once again nears the tunnel exit. It is not uncommon for traditional metering to be imposed several times on a Sunday of high traffic returning to Denver. The stoppage periods often exceed 10 min. After traffic is released, a large platoon of vehicles travels EB until it meets the back of the queue.

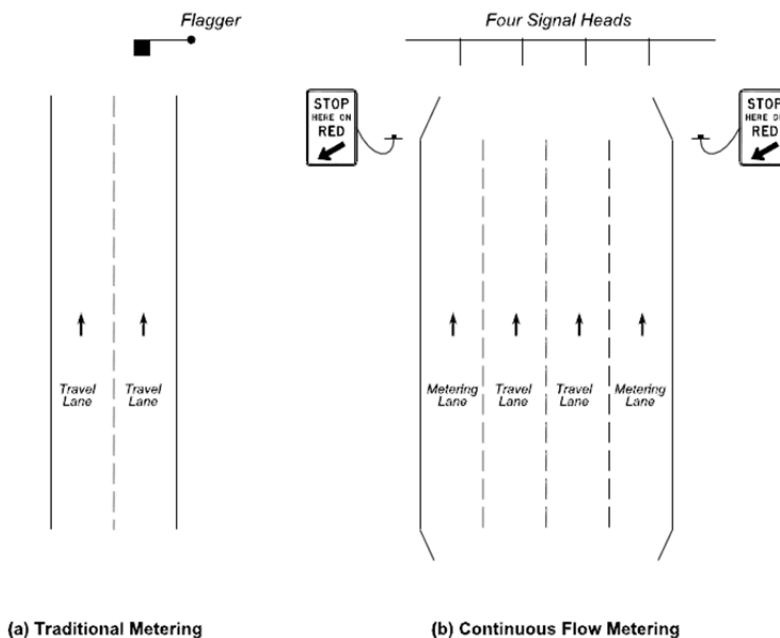


FIGURE 2 Lane configurations for the two metering schemes.

CFM regulates the flow of traffic into the tunnel similar to ramp metering. The area immediately west of the EJMT entrance was restriped to provide a four-lane metering area similar to a toll plaza approximately 50 ft wide and 500 ft long. An additional lane was striped on either side of the existing travel lanes, providing two metering areas for each travel lane as shown in the Figure 2. On an overhead gantry, four signal heads are used to indicate the stop or go for each lane. The metering lanes are released in alternating fashion (lanes 1 and 3, then lanes 2 and 4, and so on), which eliminates the need for vehicles to merge between other vehicles when entering the two tunnel lanes. Although the green release phase is set at two seconds, the frequency of releases could be manually or automatically varied to control the number of vehicles that are released per minute based on queue conditions east and west of the EJMT. After several tests, seven cycle rate plans were developed by FHU (7) for CFM implementation based on capacity and the condition and movement of the queue east of the EJMT as shown in Table 1. A 170 traffic signal controller is used for implementation due to its capabilities.

TABLE 1 CFM Cycle Rate Plans

Plan	Green Time per Phase (s)	Red Time (s)	Cycle Rate (s)	Capacity (vph)
1	2	10	24	600
2	2	8	20	720
3	2	6	16	900
4	2	5	14	1,030
5	2	4	12	1,200
6	2	2	8	1,800
7	2	1	6	2,400

DATA COLLECTED FOR THE METERING SIMULATION

Incidents and recurrent congestion occur almost every weekend and create the need to meter eastbound traffic at the tunnel. Traffic data collected for 14 high traffic days during 2010 and 2011 were reviewed in detail to select a representative day. January 30, 2011, was selected as the case study day for these simulation analyses since 1. that day had the highest EB traffic volumes for the 2010–2011 winter season, 2. incidents occurred and traditional metering was applied eight times between 1:00 and 6:00 p.m. with the longest stop duration being 25 min and the shortest stop duration being 8 min as shown in Figure 3.

Remote traffic microwave sensors (RTMS) collected traffic volumes and speeds in 2-min bins while the automatic traffic recorder (ATR) in the tunnel collected traffic volumes on an hourly basis. Colorado DOT staff stationed at the EJMT kept a manual log of all traditional metering times and durations.

Figure 4 plots the actual speed data collected on Sunday, January 30, 2011. As can be seen by this graph, severe congestion occurred between 12:00 and 7:00 p.m. The figure shows that the speed dropped at 12:40 p.m. due to an incident at Hermans Gulch (MP 218.5). To avoid the queue in the tunnel, EJMT staff first metered the traffic from 1:16 to 1:41 p.m. (25 min).

MODEL DEVELOPMENT, CALIBRATION, AND VALIDATION

Simulation is defined as “a numerical technique for conducting experiments on a digital computer, which may include stochastic characteristics, be microscopic or macroscopic in

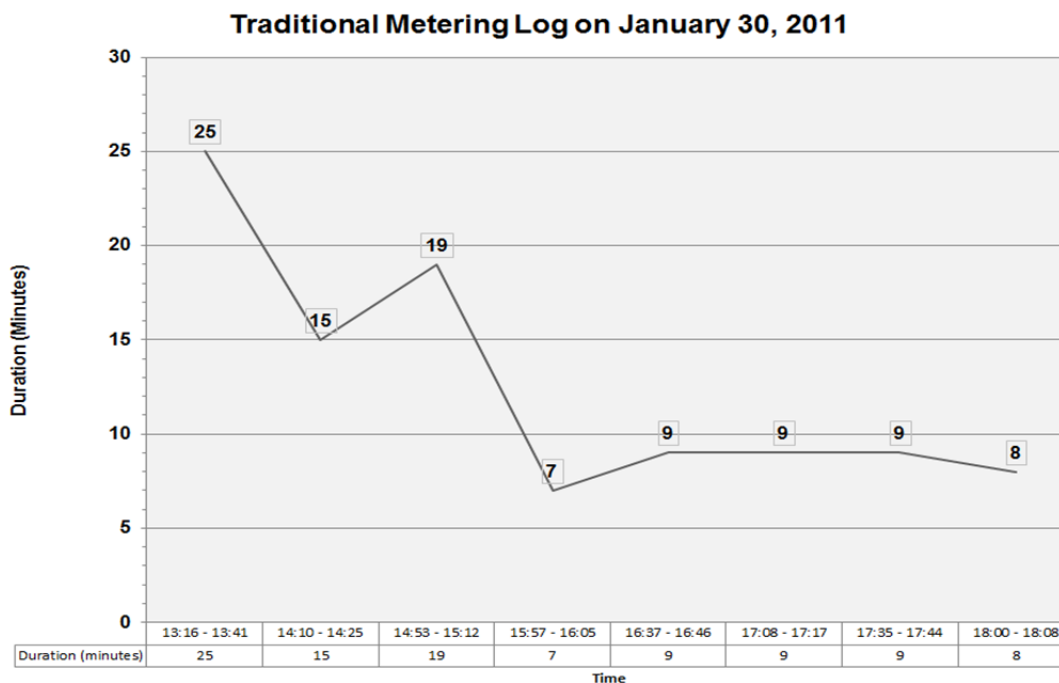


FIGURE 3 Traditional metering stoppage times on January 30, 2011.

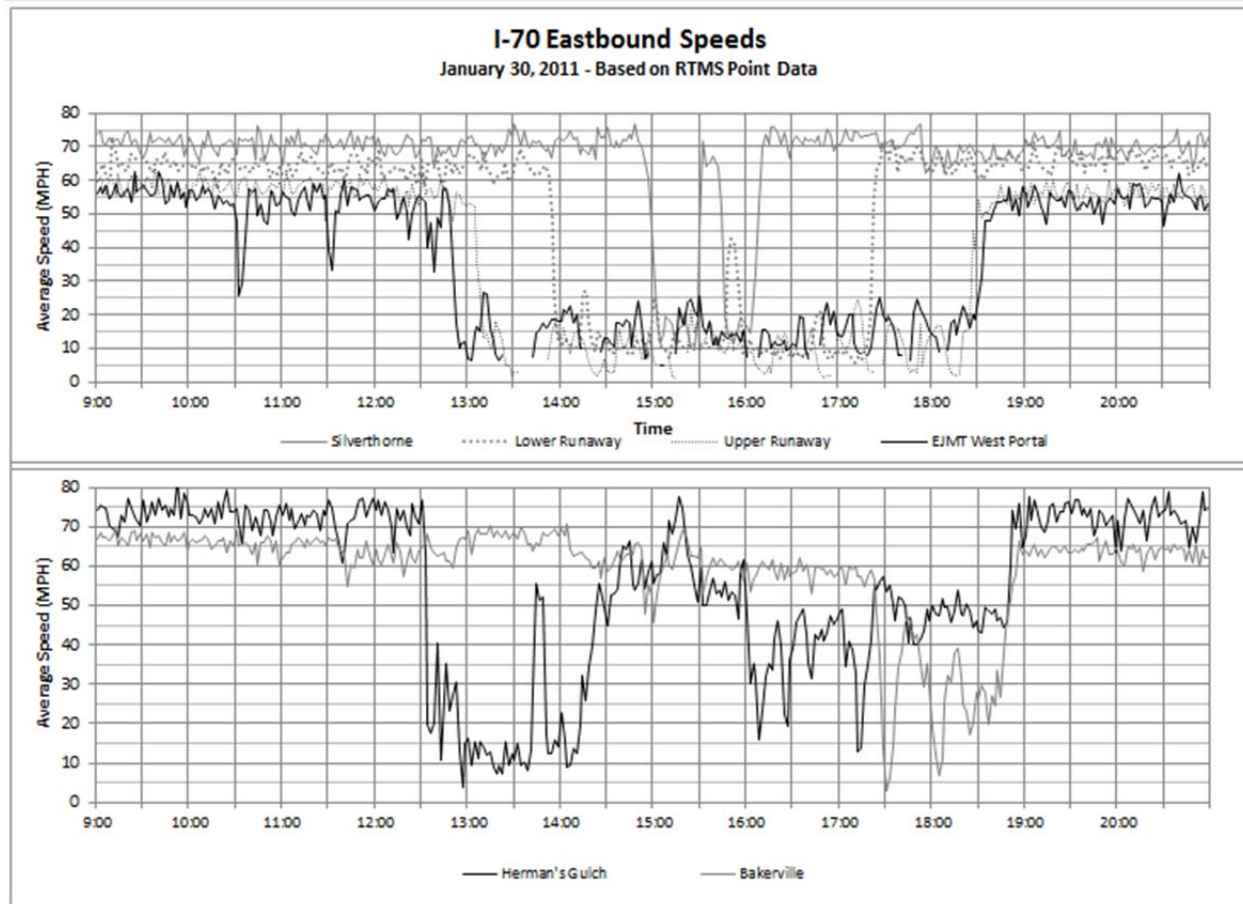


FIGURE 4 I-70 EB speeds on January 30, 2011.

nature, and involve mathematical models that describe the behavior of a transportation system over extended periods of real time” (8). The VISSIM traffic simulation model was used to simulate and quantify the traffic flow effects of these alternative metering schemes.

Using traffic simulation to replicate the actual conditions is quite challenging, because many variables such as driving behavior and highway characteristics create numerous “what if” scenarios. Developing a realistic model requires significant effort to calibrate and validate the model to replicate the observed conditions. The VISSIM model uses inputs such as vehicle lane assignments, vehicle speeds, percentages of vehicles by type, and pretimed and/or actuated signal timing (9). The experience was similar to the finding of Siuhi and Mussa (10) that “the VISSIM default parameters used in trial simulation runs for checking any coding errors showed that the default model parameters were incorrectly emulating the existing traffic flow characteristics.”

Traffic composition in the VISSIM model uses of 97% passenger cars and 3% trucks based on overall EB truck percentages on a typical Sunday afternoon compiled by Hattan and Germeroth (7). Heavy vehicles percentages are often higher during weekday business hours but less during weekend days with high recreational traffic volumes. Volume balancing was performed during all analysis time periods to ensure that volumes along the I-70 corridor were consistent between interchanges. The balanced mainline volumes were then compared to RTMS

counts, additional Colorado DOT counts, and other historical data, such as ramp volumes collected earlier by JF Sato as part of I-70 Mountain Corridor Programmatic Environmental Impact Statement. These comparisons were made to determine if traffic patterns were being replicated after the volume balancing procedure was completed. A conservative approach was adopted at locations where balanced volumes were higher than the counts and comparable historical data.

The detailed VISSIM model development for I-70 EJMT network proceeded as follows:

1. Aerial photography from Google Earth was stitched together to create the links and connectors along the corridor. Having the correct scaling is essential for this network.
2. General desired speed range distributions in free-flow conditions were used in the model per vehicle type, including passenger cars (50 to 80 mph with an average of 65 mph), trucks (40 to 60 mph with an average of 50 mph), and all vehicles through the tunnel (40 to 50 mph with an average of 45 mph). However, congestion levels and geometric conditions (grades, curves, merging, diverging, and more) that affect vehicle speed decisions were placed on every entry link and where speed limits change or on connectors with a different speed limit. Once a vehicle crossed a speed decision area, it will try to maintain that desired speed until it goes over a new speed decision point. The reduced speed zones were coded with a desired speed distribution of plus or minus 5 mph from the initial 10 mph for creating incident.
3. Lane assignments, vehicle inputs, routing decisions, and vehicle compositions were coded to represent existing conditions. Given the scale of this model keeping track of the volume through vehicle input and routing decisions at each junction is very critical to creating a realistic model. Routing decisions are carefully placed so as to not overlap and so that vehicles cannot accept new routing decisions until they have completed the previous one.
4. Driving behavior parameters, including car following and lane changing parameters, were specified as prescribed by the Wiedemann 1999 car-following model for freeways. However, due to unrealistic vehicle behavior resulting from the Wiedemann 1999 parameters, several car following and lane changing parameters [such as headway time (CC1), car following variation (CC2), and the desired safety distance] were changed based on engineering knowledge of the corridor and best practices (11).
5. Each simulation recorded from 11:00 a.m. to 9:00 p.m. ($10 \text{ h} \times 3,600 \text{ s} = 36,000 \text{ s}$) with one-hour seeding time to load the initial traffic demand.

The error checking portion of the model development focused on fixing coding errors before the calibration process began that includes a review of the coded data and a review of the animation. All coded data (geometry, speeds, signal timing data, and traffic volumes) were reviewed. The review of the animation was conducted to determine locations where general coded parameters may have been overlooked. This tuning process involved iterative parameter changing and simulating until simulated speeds closely matched speeds observed in the field. As a result, the unrealistic vehicle behavior in comparison to the field observation when using error checking also involves setting background traffic operation and driver behavior characteristics including gap acceptances, driver aggressiveness, and vehicle characteristics. The default values were adjusted during this process so that the model would accurately reflect observed conditions.

Calibration was performed to get equivalent results between the observed and expected traffic conditions (12). During validation, the VISSIM model output was compared against field data to determine if the VISSIM output was within acceptable levels. The GEH statistic

measuring the goodness-of-fit between observed and modeled flows was used for calibration purposes (13):

$$GEH = \sqrt{\frac{(E - V)^2}{(E + V) / 2}}$$

where E = model estimated volume and V = field count.

Calibration and validation were performed in three basic steps: (1) Compare bottleneck queue lengths observed in the field and in the simulation, (2) compare field traffic counts to volumes from the simulation, and (3) calibrate system performance to observed speeds and travel times.

Figure 5 shows a comparison of observed counts to simulated volumes at Silverthorne, EJMT West, Hermans Gulch, and Bakerville. When applying a stochastic model, it is not likely that there will be exact volume matches, but the volumes should be relatively close. Figure 6 shows a comparison of observed to simulated speeds at these locations plus a few more where speed data were also collected.

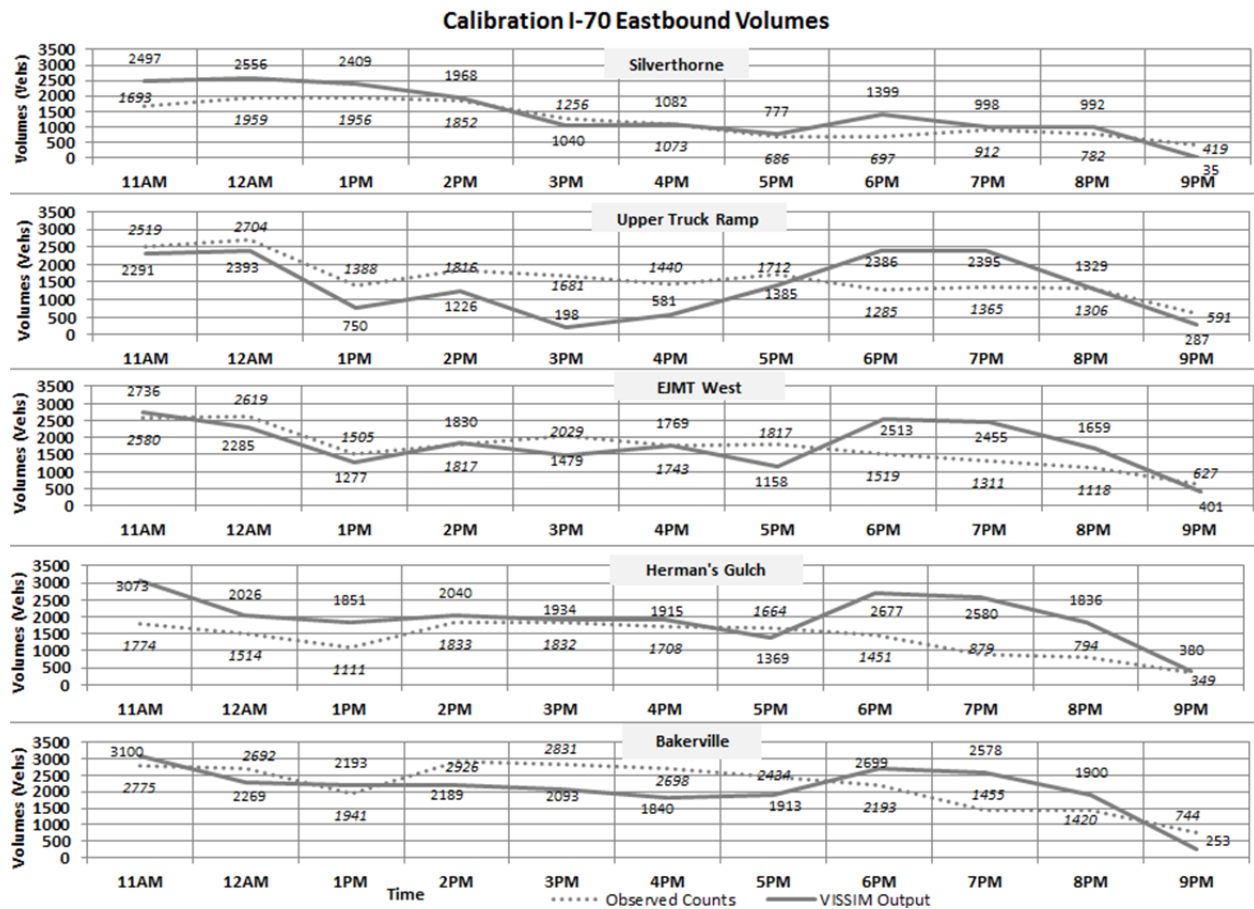


FIGURE 5 Comparison of observed versus simulated EB volumes.

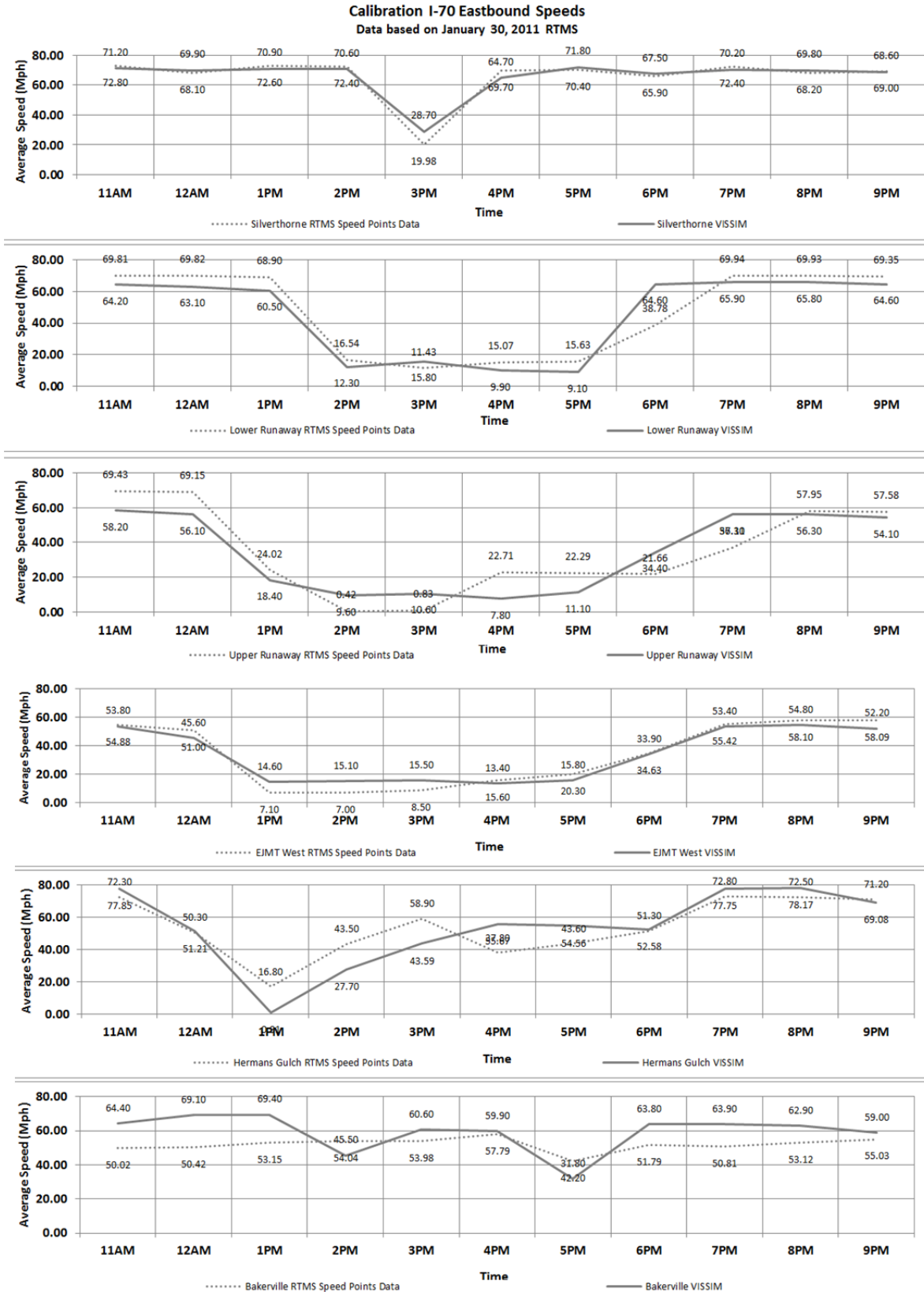


FIGURE 6 Comparison of observed versus simulated EB speeds.

Table 2 presents VISSIM output as compared to the field data and the resulting GEH statistic. GEH values less than 5 indicate a good fit, and the bottleneck location and queue length were validated with field observations. It was also confirmed through VISSIM visual animation that the traffic conditions of the model consistently coincided with field observations.

The January 30th incident was modeled in VISSIM by closing both lanes of the two-lane EB for 30 min at MP 218.5 starting at 12:34 p.m. The queue length used to validate the model was that the queue extended to Loveland Pass (about 1.0 mi) at 12:52 p.m. as observed in the EJMT staff report. The queue continued to build up to the EJMT west entrance at 1:03 p.m.

After the model was calibrated and validated using the actual data, the CFM was then applied. For the traditional metering on January 30, 2011, the average hourly travel time for the entire 16-mi analysis section from Silverthorne to Bakerville exceeded 3 h at 4:00 p.m. Table 3 compares the overall weighted average travel times and the total person-hours of travel time saved for both scenarios. Comparing the simulated queue lengths, the maximum queue length extended 7 mi using traditional metering but only 6 mi when CFM was modeled.

TABLE 2 GEH Speed Calibration Statistic

LOCATION	Description	11AM	12AM	1PM	2PM	3PM	4PM	5PM	6PM	7PM	8PM	9PM
SILVERTHORNE	Observed Points Data (V)	72.80	68.10	72.60	72.40	19.98	69.70	70.40	65.90	72.40	68.20	69.00
	VISSIM Output (E)	71.20	69.90	70.90	70.60	28.70	64.70	71.80	67.50	70.20	69.80	68.60
	GEH	0.19	0.22	0.20	0.21	1.77	0.61	0.17	0.20	0.26	0.19	0.05
LOWER RUNAWAY	Observed Points Data (V)	69.81	69.82	68.90	16.54	11.43	15.07	15.63	38.78	69.94	69.93	69.35
	VISSIM Output (E)	64.20	63.10	60.50	12.30	15.80	9.90	9.10	64.60	65.90	65.80	64.60
	GEH	0.69	0.82	1.04	1.12	1.18	1.46	1.86	3.59	0.49	0.50	0.58
UPPER RUNAWAY	Observed Points Data (V)	69.43	69.15	24.02	0.42	0.83	22.71	22.29	21.66	37.11	57.95	57.58
	VISSIM Output (E)	58.20	56.10	18.40	9.60	10.60	7.80	11.10	34.40	56.30	56.30	54.10
	GEH	1.41	1.65	1.22	4.10	4.09	3.82	2.74	2.41	2.81	0.22	0.47
EJMT WEST	Observed Points Data (V)	54.88	51.00	7.10	7.00	8.50	15.60	20.30	34.63	55.42	58.10	58.09
	VISSIM Output (E)	53.80	45.60	14.60	15.10	15.50	13.40	15.80	33.90	53.40	54.80	52.20
	GEH	0.15	0.78	2.28	2.44	2.02	0.58	1.06	0.12	0.27	0.44	0.79
HERMANS GULCH	Observed Points Data (V)	72.30	50.30	16.80	43.50	58.90	37.80	43.60	51.30	72.80	72.50	71.20
	VISSIM Output (E)	77.85	51.21	0.91	27.70	43.59	55.67	54.56	52.58	77.75	78.17	69.08
	GEH	0.64	0.13	5.34	2.65	2.14	2.61	1.56	0.18	0.57	0.65	0.25
BAKERVILLE	Observed Points Data (V)	50.02	50.42	53.15	54.04	53.98	57.79	42.20	51.79	50.81	53.12	55.03
	VISSIM Output (E)	64.40	69.10	69.40	45.50	60.60	59.90	31.80	63.80	63.90	62.90	59.00
	GEH	1.90	2.42	2.08	1.21	0.87	0.28	1.71	1.58	1.73	1.28	0.53

TABLE 3 Overall Weighted Average Travel Times

No.	Scenarios	Average EB Travel Time (min)	Person-Hours of Travel Time Saved
1.	Traditional metering	80.4	—
2.	CFM	72.2	759.0

CONTINUOUS FLOW METERING ON FEBRUARY 24, 2013

CFM was implemented during heavy traffic conditions on Sunday, February 24, 2013; it started at 11:50 a.m. and ended at 6:05 p.m. Over the course of 6 h, the CFM timing as implemented in the field was simulated using Plan 5 for a total of 24 min, Plan 6 for a total of 334 min, and Plan 7 for a total of 17 min. The traditional metering scheme was simulated for nine separate periods to replicate similar conditions when CFM was implemented. Table 4 shows the CFM and traditional metering schedules modeled.

ANALYSIS OF SIMULATION RESULTS

Three measure of effectiveness (MOEs) were selected to compare the traditional metering to CFM. These include total average travel time through the 16-mi study area (from Silverthorne to Bakerville), the total person hours of travel time, and the queue lengths west of the EJMT.

As shown in Figure 7, the simulation model estimated a total average travel time of 84 min using traditional metering on January 30, 2011, while the proposed CFM metering would have resulted in a total average travel time of 76 min for same this day. The simulation resulted in a total average travel time of 62 min using CFM on February 24, 2013, while traditional the metering was estimated to result in a total average travel time of 72 min for this same day.

TABLE 4 CFM and Traditional Metering Schedules

CFM		Traditional Metering	
11:50 a.m.–12:53 p.m.	2-s green time and 2-s red time (Plan 6) for 63 min	11:50 a.m.–12:55 p.m.	20-min full stop and 45-min release traffic
12:53–1:10 p.m.	2-s green time and 1-s red time (Plan 7) for 17 min	12:55–1:40 p.m.	15-min full stop and 30-min release traffic
1:10–1:58 p.m.	2-s green time and 2-s red time (Plan 6) for 48 min	1:40–2:20 p.m.	10-min full stop and 30-min release traffic
1:58–2:16 p.m.	2-s green time and 4-s red time (Plan 5) for 18 min	2:20–3:00 p.m.	10-min full stop and 30-min release traffic
2:16–3:01 p.m.	2-s green time and 2-s red time (Plan 6) for 45 min	3:00–3:27 p.m.	7-min full stop and 20-min release traffic
3:01–3:07 p.m.	2-s green time and 4-s red time (Plan 5) for 6 min	3:27–3:54 p.m.	7-min full stop and 20-min release traffic
3:07–6:05 p.m.	2-s green time and 2-s red time (Plan 6) for 178 min	3:54–4:19 p.m.	5-min full stop and 20-min release traffic
		4:19–4:44 p.m.	5-min full stop and 20-min release traffic
		4:44–5:09 p.m.	5-min full stop and 20-min release traffic

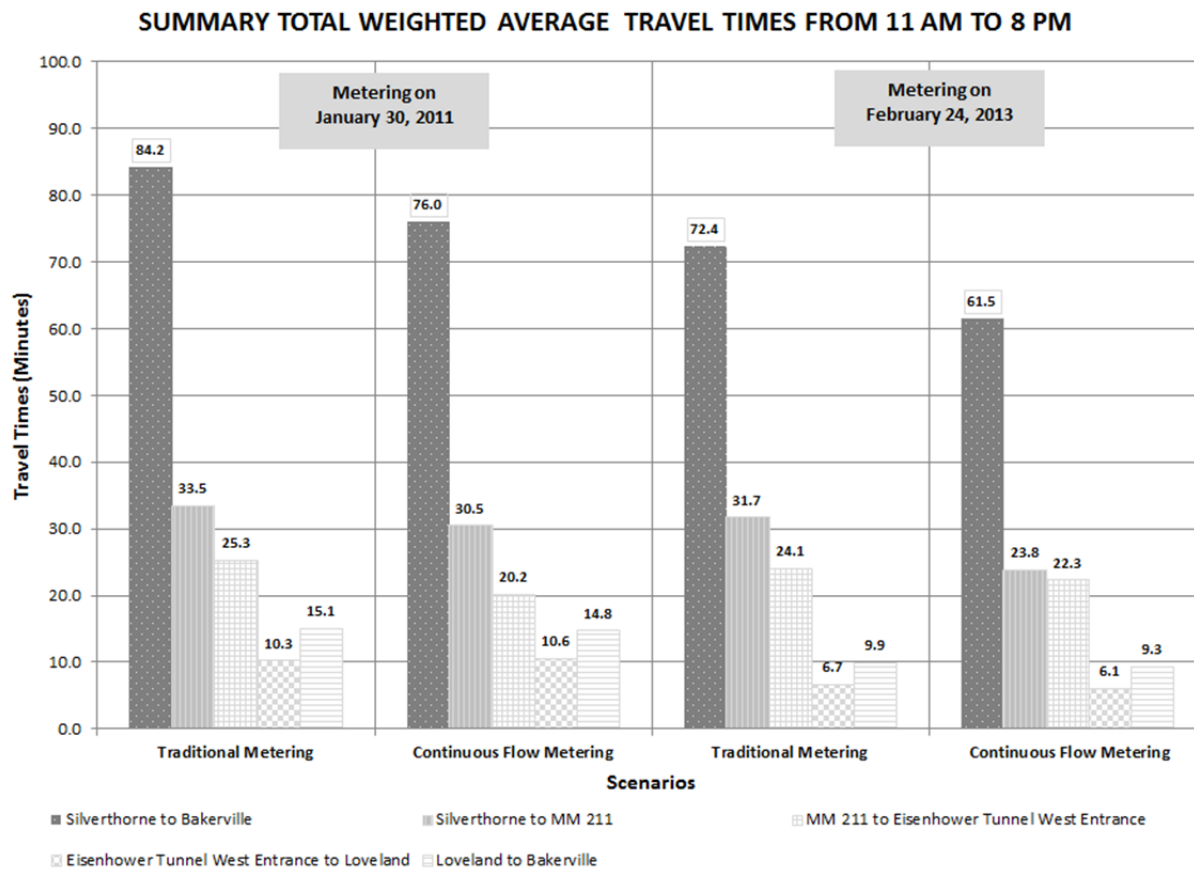


FIGURE 7 Summary weighted average travel times.

As shown in [Table 5](#), the simulation model estimated that CFM would have saved 855 person-hours of travel time on January 30, 2011, with an even greater savings of 1,134 person-hours on February 24, 2013. These person-hours of travel time assume an average vehicle occupancy of 2.6 persons per vehicle.

[Figure 8](#) shows an hourly comparison of estimated travel times for the full 16-mi segment from Silverthorne to Bakerville over the 9-h study period. With traditional metering, travel times ranged significantly from 14 min early in the day to a high of 177 min (about 3 h) for vehicles entering this segment during the hour starting at 4:00 p.m. The maximum travel time with the CFM strategy was only 120 min, which is nearly a full hour less for travelers during this hour. Maximum travel times on February 24, 2013, were estimated to have been reached at approximately 3:30 p.m.; these were 108 min using the CFM strategy and 125 min if traditional metering were used.

The third MOE examined was the maximum queue length reached on the west side of the tunnel before the metering location and propagating down the hill toward Silverthorne. It's important to reduce this queue to maintain safety and reduce fuel consumption in addition to travel time. As shown in [Figure 9](#), significant queues developed between 1:00 and 6:00 p.m. as a result of using traditional metering in the simulation model, as is observed in the field.

TABLE 5 Person-Hours of Travel Time (TT) Saved

Date	Scenarios	TT (h) of all 16-mi Trips (EB)	Person-Hours	Person-Hours of TT Saved
January 30, 2011	Traditional metering	3,368	8,757	—
	CFM	3,039	7,902	855
February 24, 2013	Traditional metering	2,460	6,396	—
	CFM	2,896	7,530	1,134

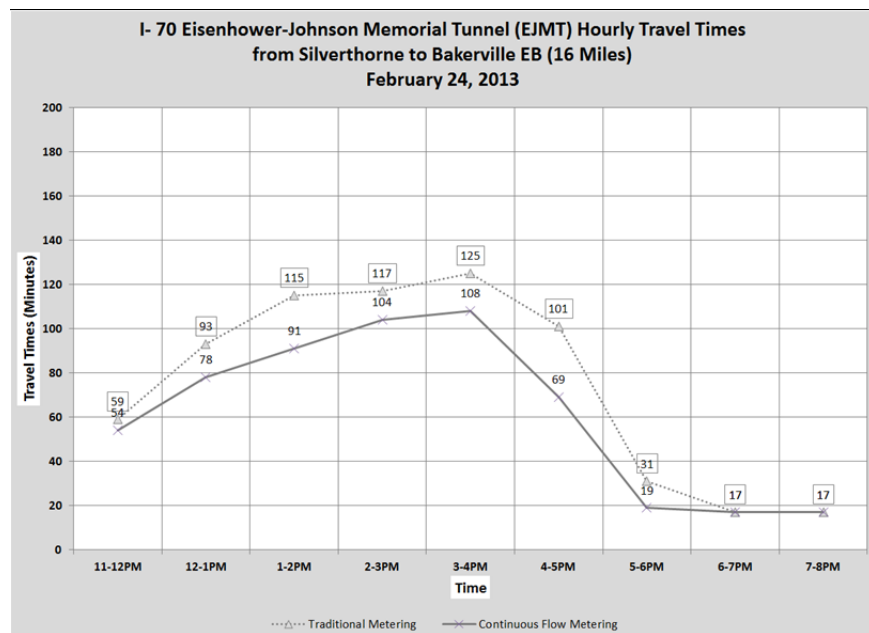
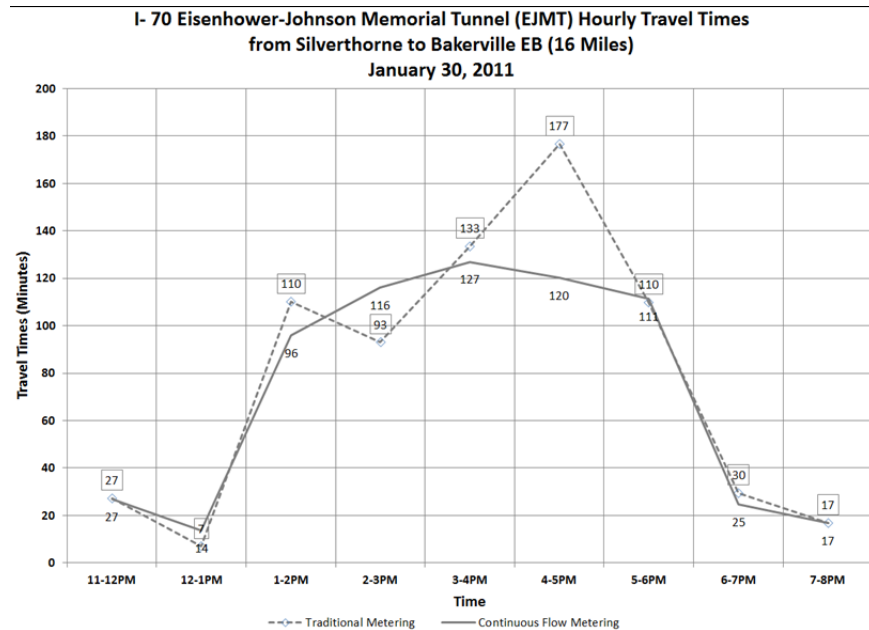


FIGURE 8 Hourly TT before and after implementation.

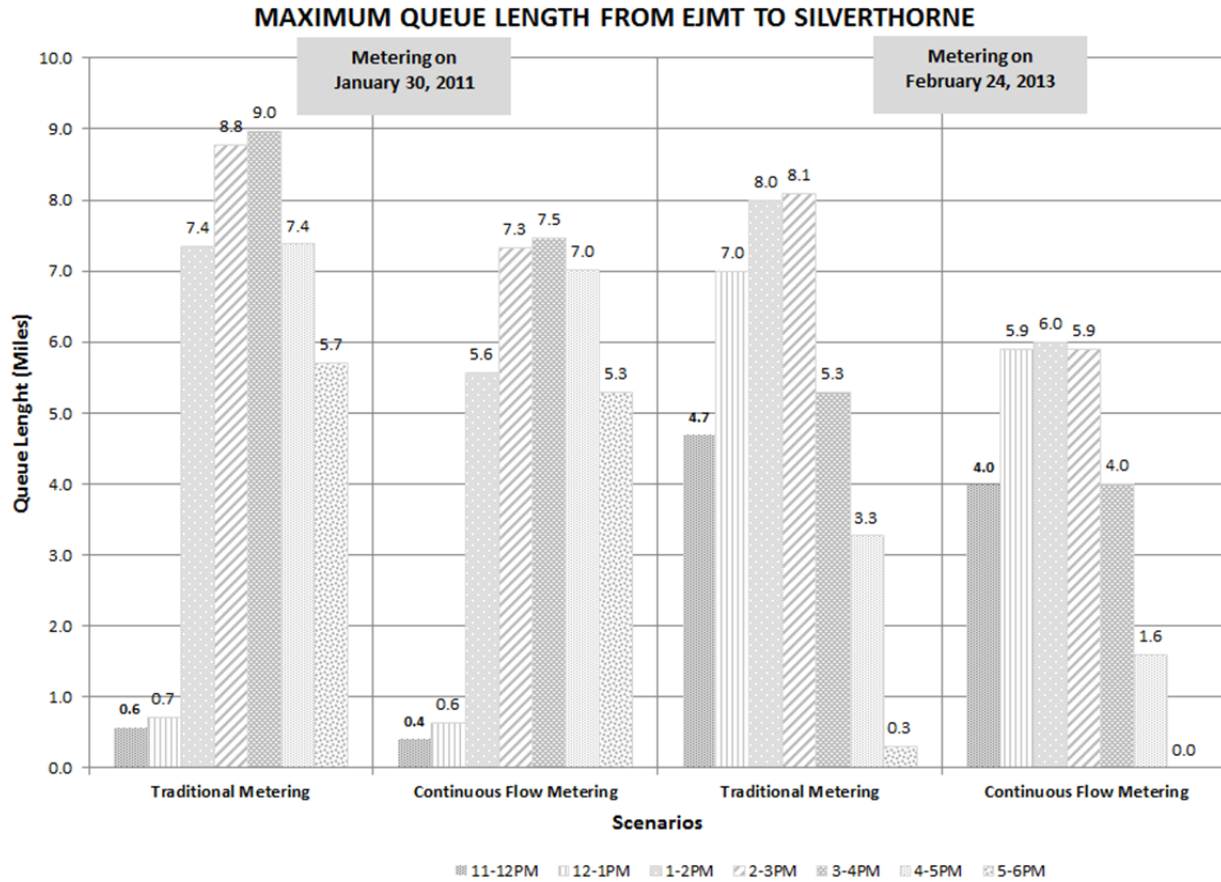


FIGURE 9 Maximum queue lengths.

The queue was estimated to reach a maximum length of 9 mi on January 30, 2011, between 3:00 and 4:00 p.m. with traditional metering, but only 7.5 mi with CFM. The queue was estimated to reach a maximum length of 8.1 mi on February 24, 2013, if traditional metering were used, but only 6.0 mi with CFM. These estimates indicate that the longest queues to develop west of the EJMT are consistently shorter with CFM than with traditional metering.

CONCLUSIONS AND LESSONS LEARNED

Simulation can be a very powerful tool for comparing alternative traffic operation plans. Of course, the simulation model requires careful construction and calibration to existing conditions regarding roadway characteristics, vehicle types, driver types, allowable travel speeds, lane changing preferences, and time-of-day variations in entering and exiting traffic volumes. There are many lessons to be learned in the process of developing and implementing a simulation model for a real-world application.

Despite the difficulty of proper model specification, once validated to observed data, the model can serve as a very useful tool for comparison purposes. While its estimates may be high or low when compared to observed data, depending on the application, the same model applied

to the same travel demand subject to operational differences can produce relatively dependable metrics for evaluation.

The objective of this study was specifically to compare the effectiveness of traditional metering versus CFM in serving traffic through a section of I-70 in Colorado that includes the EJMT. The simulation results indicate that the implementation of CFM on Sunday, February 24, 2013, improved traffic operations in all aspects evaluated as compared to traditional metering. The following performance measures resulting from the simulation model were better for CFM as compared to traditional metering:

- The average TT was reduced by 11 min, a 15% decrease.
- The highest hourly TT was reduced by 17 min, a 14% decrease.
- The average speed was increased by 22 mph, a 41% increase.
- Person-hours traveled was reduced by 1,134 h, a 15% decrease.
- The maximum queue length west of the EJMT was reduced by 2.1 mi, a 26% decrease.

CFM showed similar improvements for these MOEs when the travel demands of Sunday, January 30, 2011, were simulated, a day when traditional metering was in effect. Other lessons were learned in actual field implementation and monitoring of CFM at this site:

- Utilization of lanes 1 and 4, then lanes 2 and 3 worked better and minimized driver confusion based on the field observations. Heavy vehicles generally stay in the right lane (lane 3).
- It is necessary to estimate the capacity of the tunnel before implementing CFM, to ensure the volume of vehicles allowed to enter the tunnel does not exceed capacity.
- ATR is needed to measure volumes in the tunnel to verify the volume flowing past the metering location. RTMS and field observations are needed to monitor the queue length upstream of the metering location. TT runs are needed for actual TTs.
- Using the real-time traveler information or variable message sign (VMS) upstream of the metering area helped the drivers when metering was in effect. The two VMS messages that read “TUNNEL METERING AHEAD” and “USE ALL FOUR LANES” were used.
- The signal heads for CFM include a downward green arrow, a solid yellow one, and a solid red one. The large combined flip signs “STOP HERE ON RED” with a diagonal downward arrow and “1 VEHICLE PER GREEN IN EACH LANE” are placed to guide the driver when metering is in effect.

Further research is needed to analyze CFM using different microscopic simulation software (e.g., TransModeler or AIMSUN) in different settings with different volumes. Although this high altitude crossing of the Continental Divide is unique to Colorado, the usefulness of CFM in other settings is expected.

ACKNOWLEDGMENTS

The author thanks many people who have contributed to this study, both directly and indirectly: Dave Hattan (Felsburg Holt & Ullevig), Scott Thomas and Bart Przybyl (Apex Design), Will

Johnson (Parsons Brinckerhoff), and Kathie Haire (HDR Inc.), who provided valuable guidance and data, supportive assistance, and helpful discussions on this study, and Travis Boone and Mark Schaefer (URS) who provided travel support to present this paper.

REFERENCES

1. *Highway Capacity Manual 2010*. Transportation Research Board of the National Academies, Washington, D.C., 2010.
2. Levinson, H. S., M. Golenberg, and J. Howard. Callahan Tunnel Capacity Management. In *Transportation Research Record 1005*, TRB, National Research Council, Washington, D.C., 1985, pp. 1–10.
3. Lin F.-B., C.-W. Chang, P.-Y. Tseng, and C.-W. Su. Capacity and Other Traffic Characteristics in Taiwan's 12.9-km-Long Shea-San Tunnel. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2130*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 101–108.
4. Koshi, M., M. Kuwarara, and M. Acahane. Capacity of Sags and Tunnels on Japanese Motorways. *ITE Journal*, Vol. 62, No. 5, 1992, pp. 17–22.
5. Zhang, H. M., and W. Shen. Numerical Investigation of Stop-and-Go Traffic Patterns Upstream of Freeway Lane Drop. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2124*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 3–17.
6. Laval, J., D. Chen, K. B. Amer, A. Guin, and S. Ahn. Evolution of Oscillations in Congested Traffic, Improved Estimation Method and Additional Empirical Evidence. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2124*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 194–212.
7. Hattan, D. E., and R. Germeroth. *I-70 Eisenhower Johnson Memorial Tunnels Continuous Flow Metering (CFM) Study Draft Report*. Colorado Department of Transportation–Region 1, FHU Reference No. 10-010-18. 2013.
8. May, D. *Traffic Flow Fundamentals*. Prentice Hall, NJ, 1990.
9. *PTV Vision, VISSIM 5.40 User Manual*, PTV AG, 2012.
10. Siuhi, S., and R. Mussa. Simulation Analysis of Truck-Restricted and High-Occupancy Vehicle Lanes. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2012*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 127–133.
11. Marlina, S., and B. Janson. *Understanding the Dynamic of Truck Traffic Using VISSIM Micro-Simulation of Zipper Lane Option on Interstate 70*. ITE Western District Annual Meeting. Anchorage, Alaska, 2011.
12. *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software*. Publication No. FHWA-HRT-04-04. Federal Highway Administration, U.S. Department of Transportation, 2004.
13. *Traffic Analysis Toolbox Volume IV: Guidelines for Applying CORSIM Microsimulation Modeling Software*. Publication No. FHWA-HOP-07-079. Federal Highway Administration, U.S. Department of Transportation, 2007.

EMPIRICAL OBSERVATIONS OF TRAFFIC ANALYSIS

**Measuring the Safety Impact of
Road Infrastructure Systems on Driver Behavior**
Vehicle Instrumentation and Exploratory Analysis

JUSTIN SCHORR

SAMER H. HAMDAR

CLAIRE SILVERSTEIN

George Washington University

Featured in this pilot experimental study is the construction and design of an instrumented vehicle that is able to capture vehicle trajectory data with an extremely high level of accuracy and time resolution. Once constructed and properly instrumented, the various data collection systems were integrated with one another, and a driving experiment was conducted on Northern Virginia roadways with 18 participants taking part in the study. Trajectory data were collected for each of the drivers as they traversed a predefined loop of four roadway segments with varying number of lanes as well as shoulder widths. Data collected from the experiment were then used to calibrate the parameters of the prospect theory car-following model using a genetic algorithm calibration procedure. Once all model parameters were successfully calibrated, significance testing was carried out to determine the impacts that the varying roadway infrastructure had on driving behavior. Results indicated that there were significant changes in behavior when comparing one-lane roadways to their two-lane counterparts, specifically in cases where the roadway featured a wide shoulder. Additional testing was conducted to ensure that there was no variation based on gender as nine study participants were female and nine were male. The successfulness of this first study conducted using the newly constructed instrumented vehicle creates the opportunity for a variety of additional studies to be conducted in the future.

INTRODUCTION

Roadway infrastructure impacts driving behavior which, in turn, has significant implications when analyzing vehicle-to-vehicle interactions and assessing macroscopic transportation network performance. The main question of interest is: How does the road-surrounding environment affect the aggressive (risk attitudes) driving behavior from a traffic flow theory perspective? In order to address this question, the objective of this research is to conduct a real-world driving experiment featuring a vehicle instrumented to collect trajectory, location, and vehicle diagnostic data. Data from this experiment are then used to explicitly formulate the structure of the relationship between various car-following model parameters and one of the geometric features (shoulder width/median type) shown to be significant in previous studies (1).

Motivation and Contribution

Various methods of vehicle instrumentation have been used over the past 40 years in an effort to gain additional insights into the factors that contribute to decreased safety on roadways (2). If

total collisions are considered a surrogate measure for safety, the motivation for the examination of the different factors leading to unsafe driving conditions is highlighted by the 5,615,000 collisions that occurred on United States roadways in 2012 (an increase from the three previous years) (3). Additionally, these collisions resulted in 33,561 fatalities (an increase from the previous two years), and when considering vehicles miles traveled (VMT) as a measure of congestion the problem is exacerbated as the total VMT in 2012 was 2.969 billion, producing a fatality rate of 1.13 fatalities per 100 million VMT (both the total VMT and the fatality rate have increased over the past two years) (3). What becomes clear is that in the past couple of years roadways are trending in a direction that is both less safe and increasingly congested.

Objectives

As stated above, the main objective of this study is to understand the impact that changes in roadway geometry have on driving behavior from a traffic flow theory perspective. In order to develop this understanding, the specific objectives of this study are as follows:

- Construct an instrumented vehicle such that trajectory and headway data can be collected at a high time resolution and subsequently synced together.
- Design a real-world driving experiment using the instrumented vehicle on roadway segments with varying geometric characteristics.
- Calibrate the parameters of the prospect theory model using the data gathered from the driving experiment.
- Determine the effects that specific roadway geometric characteristics have on driving behavior through statistical analysis of calibrated model parameters.

BACKGROUND

While data driven approaches (predominately focused around the modelling and evaluation of collision data) are commonplace in the transportation research community, new and affordable technologies have led to advancements in the collection of real-time driving data. The quantification of driving behavior in real-time is an important advancement in the assessment of roadway safety, allowing for new insights through a variety of different methodologies and their subsequent applications. Three main approaches are used for the collection of real-time data: driver simulators, naturalistic studies, and instrumented vehicles, all of which have an associated set of pros and cons.

Driver simulators have been used extensively in a wide range of applications including (but not limited to) assessment of driver distraction (4), the performance of active safety and information systems (5), and the evaluation of impaired drivers (6) as well as those with certain medical conditions (7). Driver simulators are particularly useful as they allow for simulated driving experiences to be conducted in a safe and controlled environment where various scenarios (including complicated and high-risk environments) can be created and held constant for all participants in a given study (8). However, the obvious drawback to these studies is that they do not take place on actual roadways and are unable to capture the natural interactions that occur between drivers in the real-world environment (9). As such, on-road data collection methods such as naturalistic studies and instrumented vehicles are becoming increasingly

popular in order to better understand road safety crash risks and risk factors (2).

Naturalistic approaches use unobtrusive methods (typically in participants' own vehicles) to collect data in real traffic conditions (2). Again, the applications naturalistic studies are vast, including (but not limited to) examination of risks to heavy vehicle operators through the use of data acquisition systems, internal and external cameras, and daily activity registers (10); assessment of heavy vehicle operator response to a forward collision warning system through the use of gaze monitoring and brake pedal position (11); examination of older driver engagement in secondary activities at intersections through the use of a video camera system as well as a vehicle diagnostic logging system (12); and the analysis of rapid deceleration events for older drivers through the use of a custom driver monitor system that featured a two-axis accelerometer (13). Naturalistic studies allow for the collection of large amounts of data (both in terms of the number of participants and the number of trips made) over an extended period of time. Furthermore, the instruments used to collect data are unobtrusive (14), and these types of studies do not require a researcher to be present in the vehicle during data collection. The collection of this baseline data is intended to reflect normal driving (9). However, practical and analytical challenges are commonplace in naturalistic studies as datasets are large and complicated, often requiring the processing of hundreds or even thousands of hours of vehicle-based and video data (2). Additionally, since no variables are controlled by the researcher, causal conclusions cannot be drawn from naturalistic driving studies (9).

Similar to naturalistic studies, field operational tests (FOT) are long-range studies and again involve some sort of instrumentation. In these studies objective data on situation and behavior are collected through an automated process, and subjective data are usually collected manually or electronically (9). In addition to these naturalistic studies, field operational tests and driver simulator experiments, controlled on-road studies involving instrumented vehicles offer opportunities for unique data collection through the use of multiple methods (2). These controlled on-road studies are defined by their reliance on a pre-defined route in order to determine differences in performance and behavior under varying driving conditions (9). Furthermore, from a behavior perspective, field studies using instrumented vehicles are frequently regarded as the ultimate validation stage for assessing behavioral models, safety measures, and improved road infrastructure design (15). Still, the potential drawbacks of these controlled on-road studies must be mentioned, as the studies do not collect data over a long time period (2), and they require a researcher to be present in the vehicle (potentially impacting the driver's behavior) (2, 9). With that being said, these types of studies are well suited to address research questions that are independent of exposure and that use independent factors that are stable over shorter periods of time (such as age and personality) and are excellent tools in the early stages of system development and FOT design [one example of this being a situation where drivers' headway is affected, and thus the need for additional sensors (such as lidar sensors) is required] (9). Examples of studies using this type of instrumented vehicle data collection include examination of the number and nature of errors committed by drivers in distracted and undistracted states (4), analysis of the situational awareness of both novice and experienced drivers at rail crossings (16), and evaluation of an intersection violation warning prototype (17). In addition, instrumented vehicles have been used in driver training through the benchmarking of experienced drivers (Underwood, 2013).

In addition to the behavioral applications mentioned above, driver simulators, field studies, and instrumented vehicles can allow for collection of trajectory data in order to assess and calibrate car-following models. Car-following models describe the behavior of the following vehicle as a function of the lead vehicle's trajectory, allowing for estimation or prediction of the following

vehicles' trajectory in response to the actions of the lead vehicle (18). Driver simulator experiments have been conducted to evaluate car-following behavior under normal and evacuation scenarios (19), and field tests have been conducted using loop detector data to determine distance gaps under different congestion regimes (20). While these types of studies are most certainly useful in understanding car-following behavior, instrumented vehicles allow for more detailed data collection and thus have been used frequently in data collection and calibration efforts (18).

Examples of instrumented vehicles being used for data collection and the assessment of driver behavior variability in car-following include two studies by Brackstone et al. (21, 22), where headways for drivers following the instrumented vehicle were recorded in the first study and then the research was extended (in the second study) to study the factors that influence the decision-making process of car-following. While the drivers in Brackstone's studies knew they were part of an experiment, Kim et al. (23) used an instrumented vehicle equipped with an infrared sensor, a differential GPS (DGPS) inertial distance measuring instrument, a vehicle computer, and a digital video camera to measure the position, speed, and acceleration (as well as demographic information collected from the video recordings) of the following vehicles, who were unaware that they were being monitored as part of the study. In an effort to quantify driver reaction times, Ma and Andreasson (24) equipped a vehicle developed by Volvo Technologies with a GPS system, an onboard computer, two lidar sensors (facing front and rear), as well as cameras corresponding to the sensors. The study was conducted on Stockholm roadways and the follow-the-leader behaviors of random vehicles behind the instrumented vehicle were observed.

Once data from instrumented vehicles are collected, the next step in evaluating car-following models is the calibration stage. One such study was conducted by Panwai and Dia (25), who evaluated AIMSUN, PARAMICS, and VISSIM models using instrumented vehicle data collected in Stuttgart, Germany. In this case, the instrumented vehicle was equipped with radars to record the differences in speed and headway between the instrumented vehicle and the vehicle immediately in front of it (26). Similarly, Punzo and Simonelli (27) examined Newell's model, the Gipps model, an intelligent driver model, and the MITSIM model through the use of trajectory data recorded from four instrumented vehicles. Here, the four vehicles were all instrumented with GPS devices and Global Navigation Satellite System receivers (GLONASS) to record vehicle spacing data and drove in a platoon on urban and "Sextraurban" roadways in Naples, Italy (28). One final example of a study focused around car-following model calibration using data from instrumented vehicles was conducted by Soria et al. (18). Here, a Honda Pilot SUV was equipped with four wide-coverage digital cameras, a Honeywell Mobil Digital Recorder, a GPS system, and a laptop to record geographical position, speed, spacing, left-right turn signal activation, video clips, and audio recordings. The instrumented vehicle was positioned as the follower, and only the front camera was used to determine the spacing between the leader and the follower (18). The authors then used the data obtained from the instrumented vehicle to calibrate the Gipps model, the Pitt model, the MITSIM model, and the Modified Pitt model.

RESEARCH METHODOLOGY

Vehicle Instrumentation

The instrumented vehicle used for data collection in this experiment is comprised of three systems working in unison: a lidar system, a DGPS system, and an OBD (on-board diagnostics)

monitoring system. Data from all three systems are received by an in-vehicle laptop, which generates a local timestamp for synchronization purposes. A schematic for the vehicle instrumentation (overlaid on a laser scan of the actual vehicle) is provided in Figure 1; Table 1 lists the various components.

Experimental Set-Up

The driving experiment in this study allows for observation of moment-by-moment local interactions among drivers, and measures drivers’ preferred traffic measures with known attributes (gender, age, and attitude). Furthermore, experimental set-up involves testing one of the exogenous geometric factors shown to impact safety. For this pilot study, the authors have

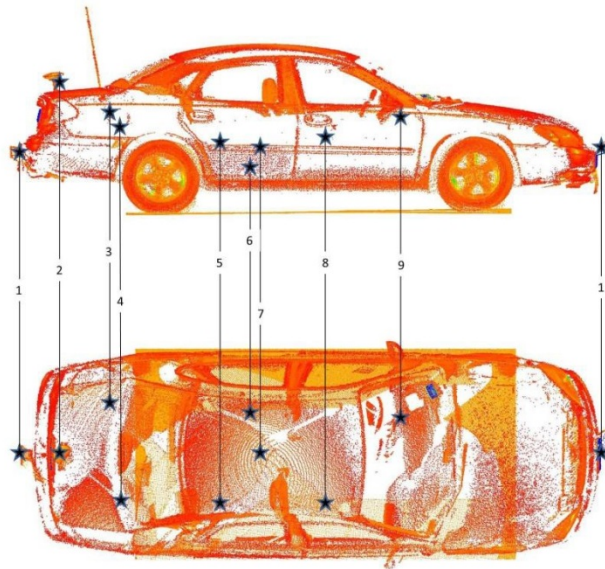


FIGURE 1 Vehicle instrumentation.

TABLE 1 Vehicle Instrumentation Key

Number	Instrument Name	Data Collected
1	Lidar sensors (2)	Trajectory data
2	DGPS antenna	Vehicle position data
3	External computing unit	
4	Sync box	
5	Ethernet switch	
6	DGPS receiver	Vehicle position data
7	Power box	
8	Laptop	
9	On-board diagnostics logger	Vehicle diagnostic data

selected shoulder width/number of lanes as the test variables and a driving experiment was conducted in an interrupted flow scenario. **Figure 2** displays a GoogleEarth image of the Northern Virginia roadway segments selected for this experiment were generated by the differential GPS data recorded during experimentation. The black line in the figure is the actual DGPS path travelled by a study participant, and the base stations zdc11910 and lwx11910 (used to increase the accuracy of the DGPS recordings) are seen in the top left and bottom center of the figure. Additionally, each of the four segments is highlighted in the figure, where the red lines mark the start and/or end point of a segment. Segment one is a two-lane roadway with a wide shoulder, Segment two is a one-lane roadway with a wide shoulder, Segment three is a two-lane roadway with a narrow shoulder, and Segment four is a one-lane roadway with a narrow shoulder. For the experiment, 18 drivers (nine males and nine females between the ages of 20 and 33) drove the instrumented vehicle through all four roadway segments. Drivers were instructed to behave as they would normally, with the exception that they were not permitted to pass the lead vehicle at any point during the test run. The lead vehicle was operated by an author of this study and speed was varied (± 7 mph from the posted speed limit) on as consistent of a basis as possible (given the surrounding traffic conditions), at the same locations throughout each of the four segments.



FIGURE 2 Roadway segments used in this pilot study.
(Image from GoogleEarth. Accessed July 23, 2014.)

Modeling and Calibration

Drivers evaluate their acceleration choice options based on the resulting potential gains and losses. Prospect theory (29) has been used to model this decision-making process (30). Here, drivers frame the stimulus where different utilities are assigned to different acceleration choices considering different weights for gains and losses and then edit the choices based on a prospect index calculated in the same way as expected utility are calculated. The prospect theory value function is formulated as follows:

$$U_{PT}(a_n) = \frac{[w_m + (1-w_m)(\tanh(\frac{a_n}{a_0}) + 1)]}{2} \left[\frac{(\frac{a_n}{a_0})}{1 + (\frac{a_n}{a_0})^2} \right]^\gamma \quad (1)$$

where U_{PT} is the acceleration value function, a_0 is the normalization parameter, $\gamma > 0$ is a sensitivity exponent indicating how sensitive a driver is toward gains or losses in travel times (i.e., speeds), and w_m is the relative weight of losses compared to the gains. Here, a driver choosing a_m as his or her desired acceleration will gain U_{PT} unless he or she is involved in a rear-end collision. Furthermore, a crash seriousness term $k(v, \Delta v)$ is used to calculate the disutility resulting from a crash as follows:

$$U(a_n) = (1-p_{n,i})U_{PT}(a_n) - p_{n,i}w_c k(v, \Delta v) \quad (2)$$

where $p_{n,i}$ is the subjective probability of being involving in a crash at the end of a car-following duration; $p_{n,i}$ is approximated by a normal distribution given that drivers are assumed to estimate the future speed $v_{n-1}(t + \Delta t)$ of vehicle $n - 1$ to be normally distributed with a mean equal to the current speed $v_{n-1}(t)$ and a standard deviation of $\alpha \times v_{n-1}(t)$ (α is a velocity uncertainty parameter); $U_{PT}(a_n)$ is derived from Equation 1, and w_c is a crash weighting function, which is lower for drivers willing to take a higher risk.

Trajectory data recorded by the instrumented vehicle (velocity, acceleration and space headway) at a resolution of 0.1 s is used to calibrate the model presented above. Since headway data was not always recorded at the same time resolution as the vehicle motion data, values were interpolated based on the change in vehicle velocity between recorded headway values. Calibration was then performed on a segment by segment basis for each driver using a genetic algorithm procedure. Following the genetic algorithm description of Hamdar and Mahmassani (31):

1. A chromosome represents a parameter set of the prospect theory model discussed above and a population consists of N_{GA} such chromosomes.
2. In each chromosome generation, the fitness of each chromosome is determined via an objective function.
3. All pairs of chromosomes are extensively generated from the current population and recombined to generate new chromosomes.
4. The crossover point where two chromosomes are combined is randomly selected.
5. Excluding the chromosome with the best fitness score, all genes (model parameters) are mutated (random variation) based on a given probability. The newly generated

chromosomes are then used in the next iteration.

6. Initially, a fixed number of generations are evaluated. The evolution is then terminated when the best-of-generation score converges from one iteration to another for a given number of generations.

RESULTS AND DISCUSSION

Calibration Results and Significance Testing

Table 2 displays the descriptive statistics for the calibration results. This includes the average and standard deviation values for the calibration parameter, velocity and space and time headways for each segment. Additionally, these descriptive statistics are provided for geometric characteristics (number of lanes and shoulder width) and gender in **Tables 3, 4, and 5**, respectively.

In order to interpret the statistical significance of the change in calibration parameters based on number of lanes, shoulder width, and gender, multiple MANOVA tests were conducted (using the SAS software). Results of the MANOVA test indicate whether or not one can reject the null hypothesis, the null hypothesis being that a certain exogenous characteristic has no statistically significant effect on the change in calibration parameters. For statistical significance and the rejection of the null hypothesis, the p -value must be less than 0.05. **Table 6** displays the MANOVA results for the effects of number of lanes, shoulder width, and gender on the calibration parameters. In addition, the effect of changing segments is included at the top of this table to demonstrate that the null hypothesis can be rejected for the change in segments. If the null hypothesis could not be rejected for the changing segments as a whole, then there would be no statistical significance of the calibration results for this study.

From the table, it is clear that a change in the number of lanes has the most statistically significant effect on the change in the calibration parameters. With this in mind, the data set was separated based on shoulder width, and a MANOVA test was again conducted for the number of lanes. These results are displayed in **Table 7**.

Here it is clear that the null hypothesis cannot be rejected when considering a change in the number of lanes on roadways with narrow shoulders, but it can be rejected for a change in the number of lanes on roadways with wide shoulders.

Finally, to ensure that there was no statistically significant difference based on gender, a final MANOVA test was carried out for each segment using gender as the dependent variable. These results (**Table 8**) demonstrate that the null hypothesis cannot be rejected based on gender for any of the segments.

TABLE 2 Descriptive Statistics for All Segments

Segment	Stat.	Vel. (m/s)	Space (m)	Head (s)	ψ	γ	W_m	W_c	T_{max}	α	β	T_{corr}	RT (s)	Vel. Error
1	Avg.	15.18	33.03	2.21	5.97	0.73	3.66	89,833	5.26	0.21	6.33	17.83	0.63	0.173
	Dev.	1.60	7.94	0.66	3.73	0.62	2.18	23,796	1.57	0.09	3.39	5.23	0.73	0.074
2	Avg.	13.99	33.09	2.41	5.40	1.09	2.83	97,944	4.83	0.11	7.08	20.39	0.36	0.100
	Dev.	1.07	13.12	1.14	4.90	0.72	1.98	16,913	2.07	0.06	2.81	4.02	0.36	0.056
3	Avg.	14.71	30.52	2.10	5.64	0.63	4.11	95,000	5.16	0.19	5.60	20.83	0.72	0.169
	Dev.	1.14	6.99	0.55	4.50	0.46	2.24	25,752	0.91	0.06	2.90	4.59	0.53	0.072
4	Avg.	15.70	29.69	1.90	4.27	0.71	3.94	100,778	5.67	0.13	6.63	20.22	0.62	0.137
	Dev.	1.50	7.46	0.48	3.91	0.58	2.46	19,283	1.72	0.06	3.03	3.81	0.47	0.059

TABLE 3 Descriptive Statistics for Number of Lanes

Lanes	Stat.	Vel. (m/s)	Space (m)	Head (s)	ψ	γ	W_m	W_c	T_{max}	α	β	T_{corr}	RT (s)	Vel. Error
1	Avg.	14.84	31.39	2.16	4.83	0.90	3.39	99,361	5.25	0.12	6.86	20.31	0.49	0.119
2	Avg.	14.95	31.77	2.15	5.81	0.68	3.88	92,417	5.21	0.20	5.96	19.33	0.68	0.171

TABLE 4 Descriptive Statistics for Shoulder Widths

Shoulder	Stat.	Vel. (m/s)	Space (m)	Head (s)	ψ	γ	W_m	W_c	T_{max}	α	β	T_{corr}	RT (s)	Vel. Error
Wide	Avg.	14.58	33.06	2.31	5.68	0.91	3.25	93,889	5.05	0.16	6.71	19.11	0.49	0.137
Narrow	Avg.	15.21	30.10	2.00	4.96	0.67	4.02	97,889	5.42	0.16	6.11	20.53	0.67	0.153

TABLE 5 Descriptive Statistics for Males and Females

Gender	Stat.	Vel. (m/s)	Space (m)	Head (s)	ψ	γ	W_m	W_c	T_{max}	α	β	T_{corr}	RT (s)	Vel. Error
Female	Avg.	15.01	27.00	1.82	5.48	0.62	3.49	94,861	5.25	0.14	6.68	20.06	0.653	0.143
Male	Avg.	14.78	36.16	2.49	5.16	0.96	3.78	96,917	5.21	0.18	6.14	19.58	0.514	0.147

TABLE 6 General MANOVA Testing

Segment			
Statistic	Value	F-Value	P-Value
Wilks' Lambda	0.484	1.84	0.0106
Pillai's Trace	0.615	1.78	0.0146
Hotelling-Lawley Trace	0.872	1.90	0.0094
Roy's Greatest Root	0.571	3.93	0.0005
Shoulder Width			
Statistic	Value	F-Value	P-Value
Wilks' Lambda	0.784	1.90	0.0684
Pillai's Trace	0.216	1.90	0.0684
Hotelling-Lawley Trace	0.276	1.90	0.0684
Roy's Greatest Root	0.276	1.90	0.0684
Lanes			
Statistic	Value	F-Value	P-Value
Wilks' Lambda	0.688	3.13	0.0036
Pillai's Trace	0.312	3.13	0.0036
Hotelling-Lawley Trace	0.454	3.13	0.0036
Roy's Greatest Root	0.454	3.13	0.0036
Gender			
Statistic	Value	F-Value	P-Value
Wilks' Lambda	0.787	1.86	0.0745
Pillai's Trace	0.213	1.86	0.0745
Hotelling-Lawley Trace	0.271	1.86	0.0745
Roy's Greatest Root	0.271	1.86	0.0745

TABLE 7 MANOVA Testing for Changing Number of Lanes Based on Shoulder Width

No Shoulder–Changing Lanes			
Statistic	Value	F-Value	P-Value
Wilks' Lambda	0.717	1.14	0.3704
Pillai's Trace	0.283	1.14	0.3704
Hotelling-Lawley Trace	0.395	1.14	0.3704
Roy's Greatest Root	0.395	1.14	0.3704
Wide Shoulder–Changing Lanes			
Statistic	Value	F-Value	P-Value
Wilks' Lambda	0.555	2.31	0.0458
Pillai's Trace	0.445	2.31	0.0458
Hotelling-Lawley Trace	0.801	2.31	0.0458
Roy's Greatest Root	0.801	2.31	0.0458

TABLE 8 MANOVA Testing Based on Gender by Segment

Segment 1–Gender			
Statistic	Value	F-Value	P-Value
Wilks' Lambda	0.364	1.56	0.2725
Pillai's Trace	0.636	1.56	0.2725
Hotelling-Lawley Trace	1.749	1.56	0.2725
Roy's Greatest Root	1.749	1.56	0.2725
Segment 2–Gender			
Statistic	Value	F-Value	P-Value
Wilks' Lambda	0.235	2.90	0.0745
Pillai's Trace	0.765	2.90	0.0745
Hotelling-Lawley Trace	3.258	2.90	0.0745
Roy's Greatest Root	3.258	2.90	0.0745
Segment 3–Gender			
Statistic	Value	F-Value	P-Value
Wilks' Lambda	0.372	1.50	0.2895
Pillai's Trace	0.628	1.50	0.2895
Hotelling-Lawley Trace	1.687	1.50	0.2895
Roy's Greatest Root	1.687	1.50	0.2895
Segment 4–Gender			
Statistic	Value	F-Value	P-Value
Wilks' Lambda	0.466	1.02	0.4940
Pillai's Trace	0.534	1.02	0.4940
Hotelling-Lawley Trace	1.148	1.02	0.4940
Roy's Greatest Root	1.148	1.02	0.4940

Discussion of Results and Parameter Explanation

Based on the significance testing conducted above, results from this pilot experimental study indicate that drivers change their behavior significantly on roadways with wide shoulders when there are a varying number of lanes. With this in mind it is important to interpret the parameter values from segments one and two (displayed above in Table 2). Interpretation of the changes in the calibration parameters between these two segments requires an explanation of the “physical meaning” for each of the parameters individually. Beginning with the gamma parameter (γ), this can be thought of as a driver's sensitivity to perceived gains and losses. That is if the value function of the Prospect Theory model generally has the form seen in Figure 3, increasing gamma would be indicative of an increase in the amplitude of the curve derived from Equation 1.

Further, the parameter w_m represents the relative weight a driver puts on losses as compared to gains. Increases in this parameter are indicative of a driver who is valuing potential risks more than that of potential gains (i.e., becoming more risk adverse). Increasing the alpha parameter is indicative of a driver being more uncertain of the leader vehicle's velocity, and the beta parameter can be thought of as the drivers' sensitivity to the surrounding environment. Increasing the beta parameter could be indicative of a number of things, including a more experienced driver or one that has become impatient. The T_{\max} parameter can be thought of as the anticipation of the driver, as increasing values indicate a driver that is thinking multiple steps ahead, and decreasing values indicate a driver who has a myopic view and is thinking about what is occurring in the moment.

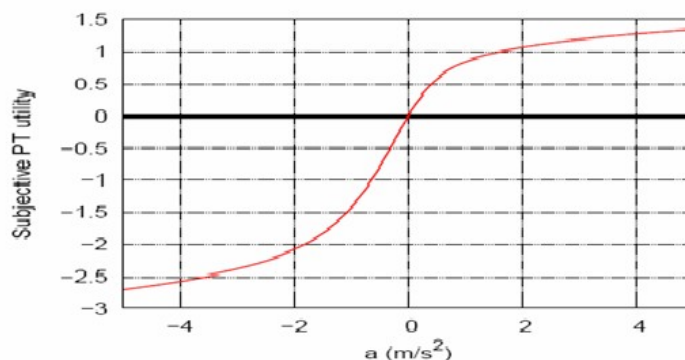


FIGURE 3 Prospect Theory value function (31).

Looking at the changes in average calibrated values for these parameters between segments one and two it appears that the one-lane segment (segment two) features higher values for beta and gamma and lower values for alpha, T_{\max} , and w_m . The combined effects of increased gamma and decreased w_m demonstrate that not only are the drivers putting less weight on perceived losses but also increasing their sensitivity to their perceived gains and losses. This result is further explained by an increase in the beta parameter, which in combination with the effects discussed above seems to indicate that drivers became increasingly impatient during this segment of the experiment. Reaffirming this notion is the decrease in the value for T_{\max} , which demonstrates that drivers are thinking more in the moment rather than anticipating what maneuvers they may make in the future (which seems to indicate a growing level of frustration). Finally, the largest percentage decrease in any parameter value is seen in that of alpha, indicating that the driver is very certain of what the vehicle in front of them is doing, once again reaffirming the notion that drivers became increasingly impatient and frustrated while traversing this segment of the experiment.

In addition to the driving environment discussed above, significance testing indicated that drivers change their behavior when moving between one- and two-lane roadways in general. The most significant changes in terms of the individual calibration parameters are seen in that of alpha, beta, and gamma. Here it is once again observed that drivers on one-lane roadways are much more certain of the lead vehicle's velocity (decreased alpha), become increasingly sensitive to their environment (or potentially increasingly impatient, increased beta), as well becoming increasingly sensitive to perceived gains and losses (increased gamma, with a slight decrease in the risk aversion parameter, w_m).

While the changes in calibration parameters were not statistically significant for shoulder width or gender, it is interesting to observe that drivers had a higher average velocity, lower space headway, and thus much lower time headway on roadways with narrow shoulders. That is, when shoulder width narrowed drivers followed much more closely to the lead vehicle. The same was true when comparing female drivers to male drivers, as female drivers had an average time headway that was nearly 0.7 s less than their male counterparts. These changes in average values were not observed when comparing one-lane to two-lane roadways, as the average velocity, spacing, and time headway were almost identical in this case.

CONCLUSIONS AND FUTURE WORK

This pilot experimental study featured the construction of an instrumented vehicle that was able to successfully capture high time resolution trajectory data through the use of multiple instruments working in unison. Furthermore, a driving experiment was successfully conducted with 18 participants driving a predefined loop that featured four segments with varying number of lanes and shoulder widths. Data collected from the driving experiment was then effectively calibrated using a genetic algorithm calibration procedure. Finally, significance testing was conducted on the calibrated parameters for the prospect theory value function, and results indicated that there were significant changes in driver behavior for varying number of lanes, specifically when the roadway featured a wide shoulder as opposed to a narrow one.

Research conducted in this study differentiated itself from that of previous studies not only with the combination of instruments that were used but also in the accuracy and time resolution of the data that were collected. Further differentiating this study from previous works, the driving experiment that was conducted tested the differences in behavior based on changing roadway geometry and then used the collected trajectory data to successfully calibrate the parameters of the prospect theory car-following model.

Given that this was the pilot experimental study for the instrumented vehicle, construction and data synchronization posed significant challenges that needed to be overcome before the actual driving experiment could take place. With these major obstacles out of the way, opportunity abounds for additional driving experiments to be conducted with a seemingly limitless potential for different types of experimental set-ups. Furthermore, the vehicle used in this study was constructed in such a manner that additional instruments can easily be integrated in the vehicle and instrumentation design, once again opening the door for a wide variety of future applications and testing.

ACKNOWLEDGMENTS

The authors thank the participants in this study. This material is based on work supported by the National Science Foundation. Any opinions, findings, and conclusions or recommendations in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. Hamdar, S., and J. Schorr. Interrupted Versus Uninterrupted Flow: A Safety Propensity Index for Driving Behavior. *Accident Analysis and Prevention*, Vol. 55, 2013, pp. 22–33.
2. Lenne, M. G., V. C. Beanland, P. M. Salmon, A. Filtner, and N. A. Stanton. Checking for Trains: An On-road Study of What Drivers Actually Do at Level Crossings. *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*, 2013, pp. 53–59.
3. NHTSA. *Traffic Safety Facts 2012: A Compilation of Motor Vehicle Crash Data from the Fatality Analysis Reporting System and the General Estimates System*.
4. Young, K., P. Salmon, and M. Cornelissen. Distraction-Induced Driving Error: An On-road Examination of the Errors Made by Distracted and Undistracted Drivers. *Accident Analysis and Prevention*, Vol. 58, 2013, pp. 218–225.

5. Liu, Y. C., and M. H. Wen. Comparison of Head-Up Display (HUD) vs. Head-Down Display (HDD): Driving Performance of Commercial Vehicle Operators in Taiwan. *International Journal of Human-Computer Studies*, Vol. 61, No. 5, 2005, pp. 679–697.
6. Akerstedt, T., B. Peters, A. Anund, A., and G. Kecklund. Impaired Alertness and Performance Driving Home from the Night Shift: A Driving Simulator Study. *Journal of Sleep Research*, Vol. 14, 2005, pp. 17–20.
7. Frittelli, C., D. Borghetti, G. Iudice, E. Bonanni, M. Maestri, G. Tognoni, L. Pasquali, and A. Ludice. Effects of Alzheimer’s Disease and Mild Cognitive Impairment on Driving Ability: A Controlled Clinical Study by Simulated Driving Test. *International Journal of Geriatric Psychiatry*, Vol. 24, 2009, pp. 232–238.
8. Bifulco, G., L. Pariota, F. Galante, and A. Fiorentino. Coupling Instrumented Vehicles and Driving Simulators: Opportunities from the DRIVE IN Project. *15th International IEEE Conference on Intelligent Transportation Systems*. Anchorage, Alaska, September 12–19, 2012.
9. Carsten, O., K. Kircher, and S. Jamson. Vehicle-Based Studies of Driving in the Real World: The Hard Truth? *Accident Analysis and Prevention*, Vol. 58, 2013, pp. 162–174.
10. Soccolich, S., M. Blanco, R. Hanowski, R. Olson, J. Morgan, F. Guo, and S. Wu. An Analysis of Driving and Working Hour on Commercial Motor Vehicle Safety Using Naturalistic Data Collection. *Accident Analysis and Prevention*, Vol. 58, 2013, pp. 249–258.
11. Wege, C., S. Will, and T. Victor. Eye Movement and Brake Reactions to Real-World Brake-Capacity Forward Collision Warnings: A Naturalistic Driving Study. *Accident Analysis and Prevention*, Vol. 58, 2013, pp. 259–270.
12. Charlton, J., M. Catchlove, M. Scully, S. Koppel, and S. Newstead. Older Driver Distraction: A Naturalistic Study of Behavior at Intersections. *Accident Analysis and Prevention*, Vol. 58, 2013, pp. 271–278.
13. Keay, L., B. Munoz, D. Duncan, D. Hahn, K. Baldwin, and K. Turano. Older Drivers and Rapid Deceleration Events: Salisbury Eye Evaluation Driving Study. *Accident Analysis and Prevention*, 58, 2013, pp. 279–285.
14. Chamadiya, B., A. Gharbi, C. Kunze, and M. Wagner. Unobtrusive In-Vehicle Biosignal Instrumentation for Advanced Driver Assistance and Active Safety. *IEEE Biomedical Engineering and Sciences*, 2010, pp. 252–256.
15. Santos, J., N. Merat, S. Mouta, K. Brookhuis, and D. D. Waard. The Interaction Between Driving and In-Vehicle Information Systems: Comparison of Results from Laboratory, Simulator and Real-World Studies. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 8, 2005, pp. 135–146.
16. Salmon, P., M. Lenne, K. Young, and G. Walker. An On-road Network Analysis-Based Approach to Studying Driver Situation Awareness at Rail-Level Crossings. *Accident Analysis and Prevention*, Vol. 58, 2013, pp. 195–205.
17. Brewer, J., J. Koopmann, and W. Najm. System Avoidance Assessment of Cooperative Intersection Collision Avoidance System for Violations (CICAS-V). *Transportation Research Record*. 2011.
18. Soria, I., L. Eleftheriadou, and A. Kondyli. Assessment of Car-Following Models by Driver Type and Under Different Traffic, Weather Conditions Using Data from an Instrumented Vehicle. *Simulation Modelling Practice and Theory*, Vol. 40, 2014, pp. 208–220.
19. Xu, Z., X. Yang, X. Zhao, and L. Li. Differences in Driving Characteristics between Normal and Emergency Situations and Model of Car-Following Behavior. *Journal of Transportation Engineering*, Vol. 138, 2012, pp. 1303–1313.
20. Dijkster, T., P. Bovy, and R. Vermijs. Car-Following under Congested Conditions: Empirical Findings. In *Transportation Research Record 1644*, TRB, National Research Council, Washington, D.C., 1998, pp. 20–28.

21. Brackstone, M., B. Sultan, and M. McDonald, M., 2002. Motorway Driver Behaviour: Studies on Car Following. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 5, 2002, pp. 31–46.
22. Brackstone, M., B. Waterson, and M. McDonald. Determinants of Following Headway in Congested Traffic. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 12, 2009, pp. 131–142.
23. Kim, T., D. Lovell, and Y. Park. Empirical Analysis of Underlying Mechanisms and Variability in Car-Following Behavior. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1999, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 170–179.
24. Ma, X., and I. Andreasson. Estimation of Driver Reaction Time from Car-Following Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1965, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 130–141.
25. Panwai, S., and H. Dia. Comparative Evaluation of Microscopic Car-Following Behavior. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 6, 2005, pp. 314–325.
26. Manstetten, D., W. Krautter, and T. Schwab. Traffic Simulation Supporting Urban Control System Development. Presented at 4th World Congress on Intelligent Transport Systems, Berlin 1997.
27. Punzo, V., and F. Simonelli. Analysis and Comparison of Microscopic Traffic Flow Models with Real Traffic Microscopic Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1934, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 53–63.
28. Punzo, V., D. Formisano, and V. Torrieri. Nonstationary Kalman Filter for Estimation of Accurate and Consistent Car-Following Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1934, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 3–12.
29. Kahneman, D., and A. Tversky. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica: Journal of the Economic Society*, Vol. 47, 1979, pp. 263–292.
30. Hamdar, S., M. Treiber, H. Mahmassani, and A. Kesting. Modeling Driver Behavior as Sequential Risk-Taking Task. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2088, Transportation Research Board of the National Academies, 2008, pp. 208–217.
31. Hamdar, S., and H. Mahmassani, H., 2009. Life in the Fast Lane. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2124, Transportation Research Board of the National Academies, 2009, pp. 89–102.

Networkwide Modeling and Control

NETWORKWIDE MODELING AND CONTROL

Toward a Systematic Exploration of the Influence of Route Choices on a Network Level of Performance

CELINE PARZANI

LUDOVIC LECLERCQ

NAZIH BENOUMECHIARA

DANIEL VILLEGAS

Université de Lyon, France

The aim of this paper is to quantify the impact of local demand distribution on the global network behavior, and, more precisely, to identify how route choices influence the level of performance of a network. A mesoscopic traffic flow simulator is chosen as an experimental platform to perform this analysis on an idealized urban center network. To simplify the exploration, route choice alternatives are clustered in homogeneous groups with respect to the percentage of overlapping. A large set of flow distributions among routes is evaluated, and different scenarios are then simulated leading to two main results. First, the fluid/congested boundary on the network is defined, and its sensitivity to parameters is evaluated. Then, route choice effects on the network level of performance are quantified using the network macroscopic fundamental diagram, which provides an aggregate vision of network behavior. Moreover, spatial distributions of traffic conditions are also investigated because heterogeneities are a well-known source of network underperformances.

INTRODUCTION

Dynamic modeling of large urban traffic networks is very challenging. Indeed, the complexity results from the combined effects of the three components of an urban traffic network: supply, demand, and traffic controls. The local supply is defined by the time-dependent capacity per link. It can be modified by traffic controls and altered by some local events. The local demand corresponds to incoming trips on a specific link. It dynamically results from the global demand (trips from origins to destinations) and the route choice process. Congestion happens when the local demand exceeds the local supply creating spatial and temporal heterogeneities at the network level.

Recent works have exhibited from empirical data in downtown Yokohama (1) a well-defined relation between space-mean flow and density. The existence of such a network macroscopic fundamental diagram (NMFD) is really appealing because it provides a simple indicator of a network level of performance. Indeed, for a given traffic state, the mean spatial speed can be known giving access to travel time estimation. If the network is homogeneously loaded, the NMFD shape will depend only on supply and traffic control. In this case, several works have been conducted leading to analytical methods for NMFD estimation taken into account for network topology and control settings (1–3). Moreover, for congested network configurations, the NMFD allows quantification of the deviation from theoretical capacity at a network level. Therefore, it allows characterizing the decrease of performance of the network

(capacity drop) caused by heterogeneities resulting either in uneven distribution of local demand or supply.

Concerning the demand impact, originally, the NMFD was found not to be influenced by the origin–destination (O-D) matrix and route choices (1). However, recent studies show that networks with an uneven (in space) or inconsistent (in time) distribution of congestion may exhibit traffic states that do not fill as a unimodal and well-defined curve. Changes seem only to be significant when the network operates near its global maximum capacity. Moreover, Geroliminis and Sun (4) show that spatial distribution of vehicle density within a reservoir significantly influences the NMFD scattering and several studies (5–7) deal with the effects of inhomogeneity of densities over a network. Recently, Knoop et al. (8) introduced the Generalized Fundamental Diagram to take into account the spatial inhomogeneity of densities that is considered as an input of the mathematical analysis. Moreover, further works on a simple parallel network have highlighted how route choice distributions modify the NMFD shape (3). In Mahmassani and Peeta (9), overall user cost and network performance under time-dependant system optimum and user equilibrium assignment patterns are examined through numerical experiments performed on a test network under various loading levels. All these works focus on the influence of network traffic distribution on the network performance estimated from the NMFD. However, few studies investigate the cause of heterogeneous network traffic states and thus make the direct connection between the network loading (distribution of the demand) and the network performance.

This paper will focus on the impact of local demand distributions on the network traffic conditions in order to improve the description of heterogeneities resulting from the network structure and the global demand profile. More precisely, the interest is in identifying how route choices influence the level of performance of the network. The aim is then to establish a methodology to characterize route choice impacts on the NMFD and on the global network performance indicators. In order to better identify which component of route choices has the larger impact, route selection will be studied apart from flow distribution. Since a fine meshed network leads to a large number of potential route choices, a clustering method to reduce the problem size by defining homogeneous sets of route selections with identical properties will be proposed.

As route choices are difficult to observe in practice for dense urban networks and are thus hard to capture, this paper follows a simulation-based approach to carry out the analysis. The outline of the paper follows. In Section 2, the global framework of the study is presented, detailing the network and the analysis designs. The proposed clustering method for route sets generation is also presented. Then, Sections 3 and 4 are concerned with the results of the simulations of different scenarios. The network level of performance is examined at different scales using several indicators.

First, the fluid–congested boundary and its sensitivity to parameters are analysed. Second, the impact of route choices is studied more in detail through the NMFD and the potential exhibited capacity drop. Then, spatial heterogeneities are quantified using standard deviation of macroscopic traffic conditions. Finally, Section 4 presents conclusions and an outlook on further researches.

NETWORK DESIGN AND ROUTE CHOICE MODELING

This study considers an idealized network mimicking an urban center. Heterogeneities are studied both analytically and by simulation. The simulations are performed using the even-based mesoscopic model proposed by Leclercq and Becarie (10), which is fully consistent with the Lighthill-Whitham-Richards (LWR) model (11, 12) at a macroscopic scale for a single-class triangular fundamental diagram. Moreover, this model provides passing times for all vehicles at all link boundaries so that vehicle trajectories within all links can be estimated. In order to complete the analysis design, supply and demand are defined in the following.

Idealized Network Design and Supply Description

The network is represented by a lattice-like road network with 12 origins and 12 destinations. Only north and west to south and east directions are considered to reduce the problem size. The 84 links of the network represent one-way road sections (see the flow directions represented by arrows on Figure 1) while the 60 nodes represent intersections. Each road section i has the same length $l_i = 300$ m. The fundamental diagram parameters are a free-flow speed $v = 50$ km/h, a jam density $k_{\text{jam}} = 200$ veh/km, and a capacity $C = 0.3676$ veh/s. Intersections are controlled by two-phase traffic signals with a fixed cycle equal to 1 mn, identical for green and red times with no offset. Note that some symmetry properties are inherent to this network design. Moreover, the network is not loaded at the beginning of simulations.

Demand Description and Route Choices Modeling

Global Demand

As described above, the studied network is composed of a set of 144 O-Ds, and for each O-D pair a uniform distribution through destinations occurs. Two parameters are then introduced to define the demand: the input flow Q_{in} and $\alpha \in [0,1]$, which allows to increment the demand value from Q_{in} to $Q_{\text{in}}(1 + 5\alpha)$ along the origins links as plotted in Figure 1. In order to ensure that the origins links are not saturated, the following is imposed:

$$Q_{\text{in}}(1 + 5\alpha) < C \quad (1)$$

where C is the link capacity.

Local Demand

The next step is the route choice description. It is well known that predicting which route a given traveller would take when going from one origin to one destination is a key step in traffic forecasting models. The route choice depends, on the one hand, on the attributes of the available routes, such as travel time, number of traffic lights, and more. On the other hand, characteristics and preferences of the traveller also influence the choice. All these aspects of the route choices problem make it particularly complex, especially for large networks representing dense urban areas. Usually, two approaches can be considered for route choices description. They can either be provided by a model based on equilibrium principle (user equilibrium or system optimum) or

by using an implicit/explicit expression for choices as for dynamic traffic assignment models. Reviews of route choice models are available in the literature in Bovy and Stern (13), in Ben-Akiva and Bierlaire (14) for discrete choice methods, and in Szeto and Wong (15) for traffic assignment. This paper will not question the equilibrium principle or traffic assignment models but instead define a framework to estimate the impact of a high variety of potential choices. Thus, a two-step process will define route choices: 1. paths selection and 2. flow distribution over paths.

A total of 3,430 potential paths are available to link the set of 144 O-Ds composing the studied network; some of them are explicitly represented in Figure 1. First note that resulting from network design assumptions, all these paths have the same length. As a consequence, the free-flow travel time is identical for all paths so that they can be clustered according to the level of overlapping. Moreover, in simulations, routes are assigned to vehicles once they enter the network.

Step 1: Clustering Method for Paths Description

A crucial property when considering the route set is the correlation between the different alternatives. Several models have been proposed to take this component into account: Cascetta et al. (16) for the C-logit model, Ben-Akiva and Bierlaire (14) for the path-size logit (PSL) model, and Vovsha and Bekhor (17) for the link-nested logit model. In this study, the route set definition was derived from the C-logit model. This model explicitly addresses the correlation among alternatives (like the two available routes from O_8 to D_8 in Figure 1). The basic idea is to treat the interdependency of the routes through a commonality factor. Thus, highly overlapped paths have

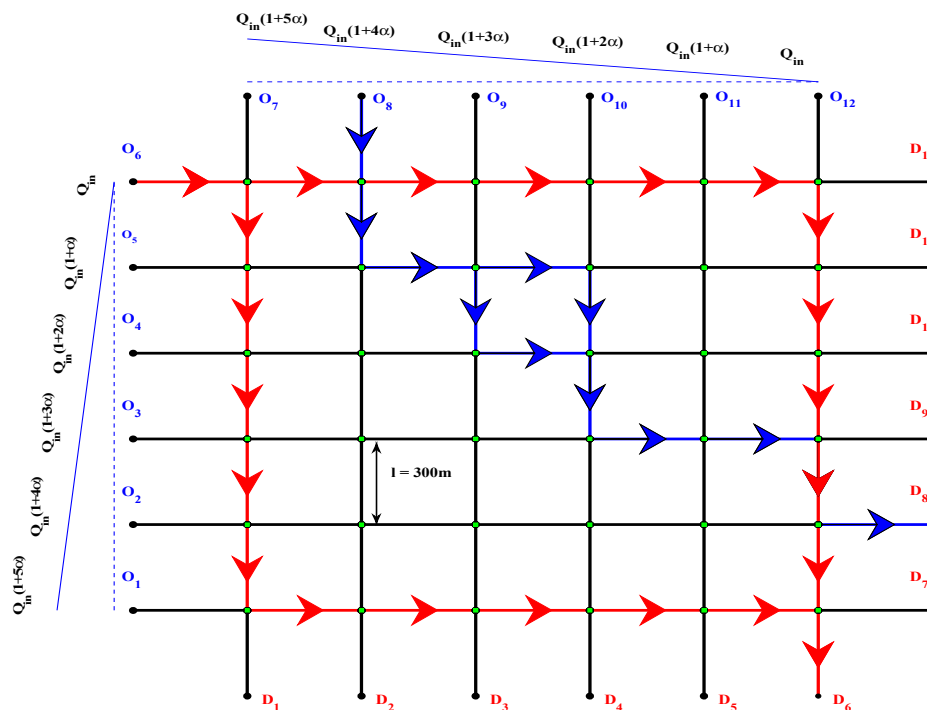


FIGURE 1 Simulations are performed with an urban center represented by a lattice-like unidirectional road network. The arrows represent flow directions.

a larger factor and therefore smaller utility with respect to similar paths. For each alternative path p_k of a given O-D pair, the common factor CF_k is proportional to the degree of overlapping of path p_k with other alternative paths. It is computed as follows:

$$CF_k = \beta \ln \sum_{l \in I} \left(\frac{L_{lk}}{L_l^{\frac{1}{\beta}} L_k^{\frac{1}{\beta}}} \right)^\gamma \quad (2)$$

where L_{lk} is the length of sections shared by paths l and k , while L_l and L_k are the length of paths l and k , respectively. Note CF_k that defines the proximity of alternatives independently of the flux on each road, which is appealing to study fluid–congested boundary dissociating the two states. In this work, it is assumed that $\beta = 1 = \gamma$ and computed CF_k for each path k of a given O-D pair. **Figure 2** shows the results for two distinct O-D pairs. In order to develop a clustering method for route set generation, it was decided to target five clusters of paths characterized by their CF_k value as plotted in **Figure 2**. Therefore, for a given O-D pair, clusters definition then corresponds to a uniform splitting of the interval defined by the minimum and maximum values of CF_k . More precisely the G_1 cluster contains two paths with no sections shared and for $i = 2$ to 5, cluster G_i contains alternative paths with more and more shared sections. Note that the distribution of CF_k is specific for each O-D as illustrated in **Figure 2**. Note also that under these assumptions, two clusters have specific characteristics: G_1 only contains the two most independant paths, while G_5 contains those which are the most correlated.

Step 2: Sample Choice for Each Cluster of Paths

Because of the definition of paths clusters, several routes are available in a given one. In order to perform simulations, a maximum of five roads for each O-D pair is imposed. For each cluster G_i , the selected paths are denoted from R_1 to R_5 , with decreasing values of the commonality factor, R_1 corresponding to the maximum. The selection method is illustrated in **Figure 2**: For each O-D pair, CF_k values for each alternative path p_k is plotted in blue, the five clusters boundary is also plotted, while the selected paths R_i in each cluster are drawn in red. The impact of this particular sampling method will be further questioned.

Once the paths are assigned to each O-D pair, the following step consists in distributing the inflow of vehicles across the five potential paths. The aim of this study is to evaluate a large variability of such a path flow distribution. For this, we consider a discrete approximation of a normal distribution depending on two parameters μ and σ . This allows for instance to simulate a uniform distribution over all paths or on the opposite to favor one of the five potential paths setting the maximum on it.

To summarize the design analysis and the demand definition, **Table 1** presents the parameters involved in this analysis and the range of values explored in the different scenarios.

In the next sections, different scenarios are analyzed using analytic and numerical computations. Two distinct analyses are performed to quantify the network level of performance. First, the fluid–congested boundary for the network and its sensitivity to parameters is presented in **Section 3**. This allows characterizing the global network behavior from a rough and binary point of view. Indeed, the focus is on dynamic loadings. When the local demand on each link does not exceed its capacity, traffic dynamics only correspond to the queue evolutions within links. No spillbacks occur and the whole network remains in free-flow conditions. Otherwise,

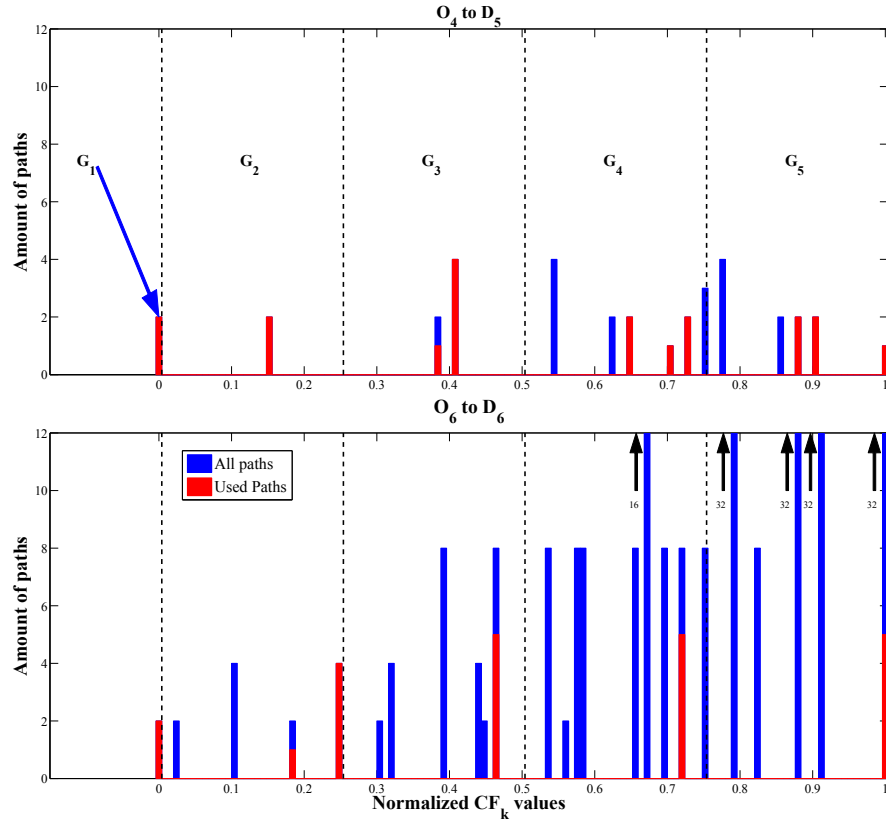


FIGURE 2 Histogram of normalized commonality factors for $\beta = 1 = \gamma$ for different O-D pairs. In red, CF_k values correspond to selected paths.

congestion will appear in any link where the local demand exceeds capacity and will propagate within the network. The network can then be considered as congested or at least partly congested.

Then, once congestion has appeared, in order to better evaluate the impact at the network level, two complementary studies are presented. In Section 4, route choices effect is investigated more in detail through numerical simulations given access to the NMFD and the potential exhibited capacity drop. Last, spatial heterogeneities are quantified by studying the distribution of traffic conditions.

TABLE 1 Parameters of the Study

α	Q_{in}	G_i	p	σ
[0,1]	$[0, Q_{max}^\alpha]$	$i \in 1, 2, 3, 4, 5$	[0, 1]	[0, 01; 10]

ANALYSIS OF THE NETWORK FLUID–CONGESTED BOUNDARY

Methodology

A simple way to characterize the network performance is to determine the binary state of the network (free-flow or congested). Interestingly, this can be done without resorting to simulations.

Indeed, the network fluid/congested boundary can be analytically computed following the mathematical formalism described in Bierlaire (18) that is briefly recalled. Refer to Section 2 for a detailed description of the studied network. Denote the O-D flows associated to each of the 144 considered O-D pairs by q and the link flows for each link by u the link flows for each link. With the notations already introduced, each O-D pair k is linked by the set of p_k alternatives paths. The link-path incidence matrix L , only depends on the network topology and is defined by $L_{kl} = 1$ if link k belongs to path l , 0 otherwise. Thus, the problem can be rewritten under the matrix form

$$LCq = u \quad (3)$$

Solving System 3 allows determining with low computational cost the network fluid–congested boundary. Indeed, as explained in Section 2, a simulation input is defined by the n -uplet of parameters $(\alpha, Q_{in}, G_i, p, \sigma)$ where α and Q_{in} are related to the demand, G_i corresponds to the cluster of paths for a given O-D while p and σ are related to the flow distribution over paths for each O-D pairs. Exploring the range of each parameter, System 3 is solved. If the results are such that there is at least one saturated link $u_j \leq C$, C being the link capacity, the n -uplet is assigned as congested. Indeed, if a congestion appears at some point, it will naturally spillback over at least a part of the network.

Results

Following this methodology, several tests are performed to study the fluid/congested boundary as network performance indicator. Figure 3 shows the evolution of the fluid/congested boundary depending on the demand: Q_{in} (x -axis) versus α (y -axis), for three values of σ : a small one, $\sigma = 0.1$, meaning that the maximal inflow is mainly concentrated on one path, a large one $\sigma = 9.1$ simulating the case of a uniform inflow distribution and an intermediate one $\sigma = 0.5$. For this test case, the central path described through the parameter p corresponds to the path R_3 denoted for each cluster. Note that when σ is large, p has no influence. The range of values for (α, Q_{in}) corresponding to the fluid network configuration respectively congested configuration is plotted in black, respectively in grey. Recall that Q_{in} parameter is constrained by the value of α as detailed in Section 1 Equation 1. Therefore, the range of values cannot be entirely explored leading to this specific shape.

Several conclusions can be deduced from Figure 3. First of all, it has to be noted that the shape of the fluid–congested boundary does not depend on either clusters' identification or σ value. Next, for a given clusters' identification (that is for a given line), comparing the three columns, the impact of σ parameter can be studied. It is clear that the higher σ is (that is the most the flow is distributed over the selected paths for the given cluster) the lower the network is congested. For instance, for Cluster G_2 , the percentage corresponding to fluid network configuration grows from 40% for $\sigma = 0.1$ to 56% for $\sigma = 9.1$. Then, for a given σ (that is for

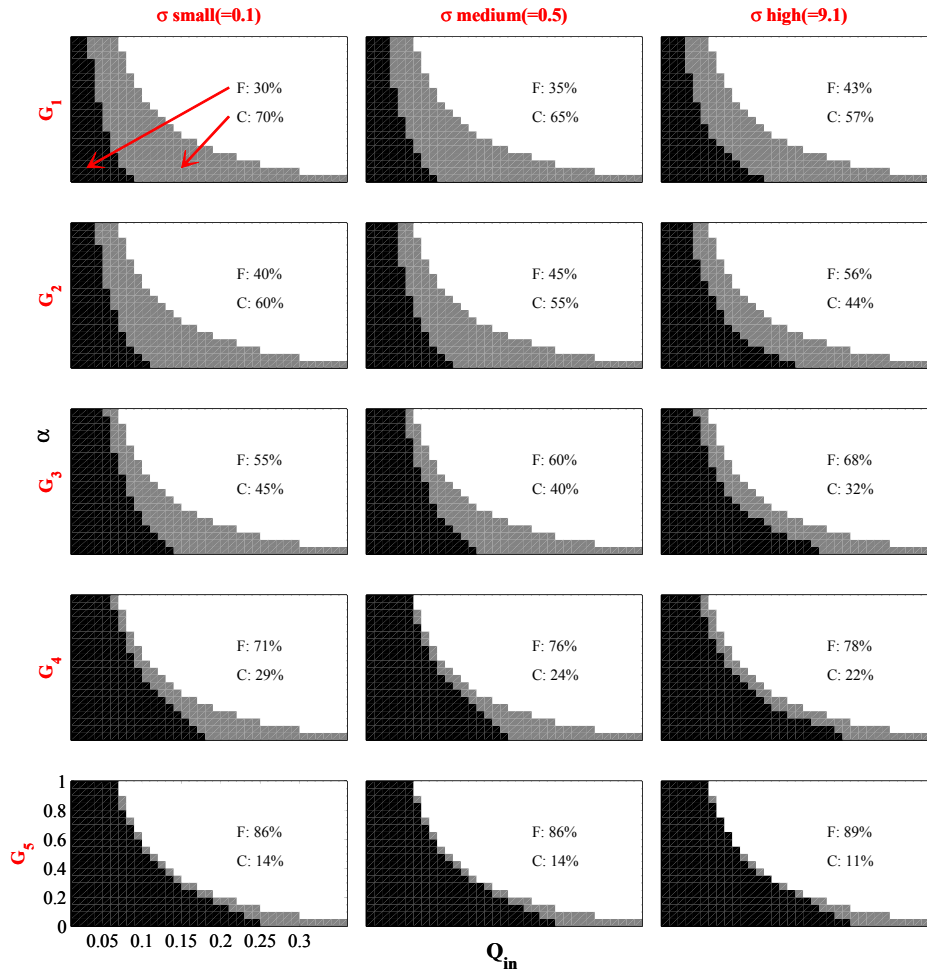


FIGURE 3 Fluid–congested boundary evolution depending on the demand. In columns, different values of σ and in lines, the level of overlapping from cluster G_1 to G_5 .

a given column), the clusters’ID analysis can be performed. Results on Figure 3 show that the higher the value of CF_k is (that is for a high clusters’ identification), the larger is the fluid zone. For instance, for $\sigma = 0.1$, the percentage corresponding to fluid network configuration grows from 30% for cluster G_1 to 86% for cluster G_5 . That means that the more the paths overlap between each O-D pair when CF_k value is high, the lower the network is congested. This is a nonintuitive conclusion since usually when considering a single O-D pair low overlapping intuitively means a better (in the sense of uniformly) distribution over the different routes. Indeed, higher overlapping values will increase the shared portions of routes with higher flows and thus a higher probability for congestion. However, the interest here is in the whole network performance and all O-D pairs need to be considered. That is why in order to improve the analysis, what the clustering method using the CF_k coefficient means at the network level will be studied in more detail. For each cluster of paths and for each link of the network, Figure 4 plots the number of paths crossing over this link normalized versus the total number of paths of each cluster. It can be observed for instance that as a result of the particular structure of the studied

network a low commonality factor value for each O-D pair (see first subplot) favors the peripheral links. Then, comparing the subplots in Figure 4, the level of global overlapping is increasing from cluster G_1 to cluster G_5 and it can clearly be observed that considering close alternatives for routes at the single O-D level may paradoxically lead to a whole set of routes that weakly overlap. Figure 5 clearly shows that the lowest CF_k value is (see cluster G_1) means that more than 50% of the network links carry very few paths (less than 10). For higher CF_k the links carry more paths with a more homogeneous distribution. This means that the repartition of paths over the network is more well balanced when CF_k inverse.

In the second test, the value of σ is assumed to be high enough to have a uniform flow distribution over the paths and no effect of the p -value. Computations are performed for different numbers of paths: 1, 2, or 5 paths in each cluster, again chosen from the maximal CF_k value in each cluster. The fluid–congested boundary depending on the demand parameters is plotted in Figure 6 for clusters G_2 to G_4 which are less specific. The main conclusion is that the higher the number of paths there are, the lower congestion can be observed. However, the most noticeable gain is obtained when the number of paths increases from 1 to 2.

When this number is higher than 5 no changes can be identified. Last, a test to study the impact of the choice for the sample of five paths in a given cluster is performed. Different options are tried, for instance five paths with an increasing CF_k value for each cluster, five paths equally distributed among a cluster. The results obtained show that there is no significant effect.

Combining the results of these tests, this study shows that the more homogeneously the flow is distributed over the network, the better the network performance. In this study, homogeneity conditions result both from high level of overlapping and a uniform flow distribution through the different paths.

REFINED ANALYSIS OF THE LOCAL DEMAND DISTRIBUTION ON THE NETWORK PERFORMANCE

In this section, the paper resorts to refined indicators to assess the network performance. Some are inspired by the concept of the NMF (e.g., network traffic evolution, capacity drop), others correspond to classical representations of heterogeneities (e.g., density standard deviation).

Different Indicators to Assess the Network Performance

Using Edie's generalized definitions (19) and assuming all vehicles trajectories are available for each link i the average flow Q_i and density K_i in the time step $[t + \Delta t]$ are given by

$$Q_i = \frac{\sum_k d_k}{l_i \Delta t} \quad K_i = \frac{\sum_k t_k}{l_i \Delta t} \quad (4)$$

where d_k is the distance traveled by vehicle k in the considered zone (corresponding to link i and time interval $[t + \Delta t]$) and t_k respectively, its time spent in the zone. Note that Δt should be of the order of several signal timing to provide accurate mean estimations. These definitions allow

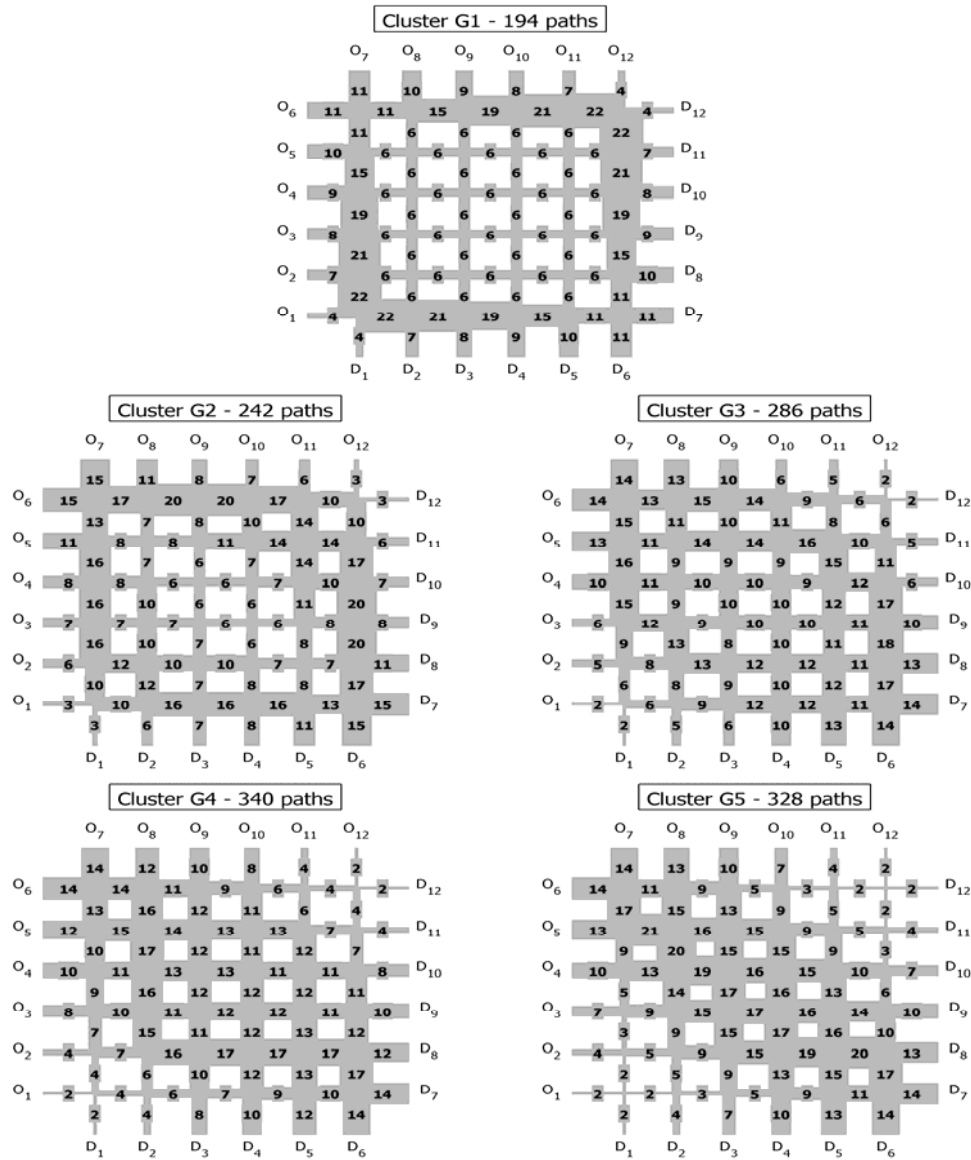


FIGURE 4 Percentage of paths crossing over each link of the network for clusters G_1 to G_5 .

determining two indicators series: on the one hand, the average network flow Q , density K , and speed V defined as follows:

$$Q = \frac{\sum_i l_i Q_i}{\sum_i l_i} \quad K = \frac{\sum_i l_i K_i}{\sum_i l_i} \quad V = \frac{Q}{K} \tag{5}$$

The time evolution of Q and K provide a synthetic vision of the network global behavior. For simplicity, (Q, K) plots are provided for successive time periods.

On the other hand, the standard deviation of the density across the network is used as an indicator of spatial distribution of network conditions.

Moreover, simulations duration is chosen large enough in order to ensure that the network reaches in a steady state. This has been monitored by computing the cumulative curves

for the total in- and outflows. When the difference between these curves (that corresponds to the total accumulation within the network) is constant, the network is considered to be in a steady state. Note that when the network is congested the difference can never be constant because oscillatory patterns are observed. In that case, the network is considered to be in a quasi-steady state when the magnitude of the variations is small enough.

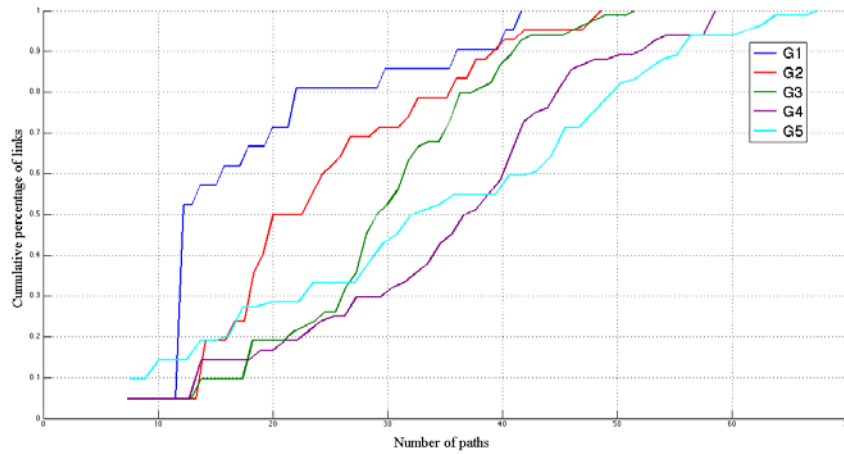


FIGURE 5 Percentage of links that carry at least x paths (repartition function).

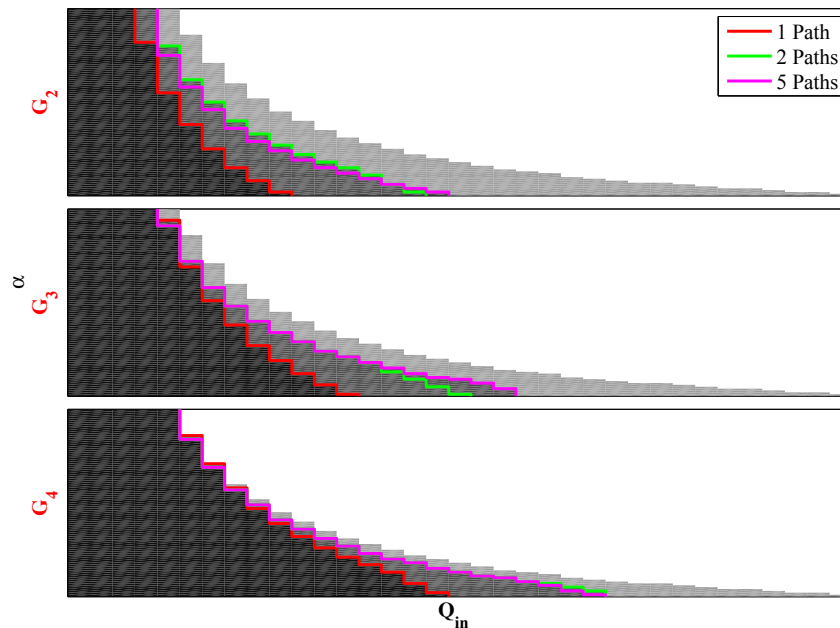


FIGURE 6 Fluid-congested boundary evolution for different number of paths with uniform inflow distribution for different levels of overlapping, clusters G_2 to G_4 .

Results

The first test aims at evaluating the (Q, K) plots sensitivity to the level of overlapping representing through the paths clusters in this analysis. Two cases are plotted in Figure 7 for a fixed demand corresponding to $\alpha = 0.15 = Q_{in}$. In the upper subplot (a) $\sigma = 10$ corresponds to a uniform distribution whereas in the bottom subplot (b) $\sigma = 0.1$. Figure 7 clearly shows that the NMF shape depends on the degree of overlapping of paths: for instance on (a) for clusters G_3 to G_5 , the network remains in free-flow condition, whereas for clusters G_1 and G_2 it switches to congestion. Comparing the two plots, it can be observed the impact of σ since cluster G_3 now experiments congestion in (b). Moreover, these plots give access to an estimation of the network capacity drop represented by the difference between solid and dot lines for each cluster.

The reference for capacity is assumed to be the maximal obtained with one of the clusters for a given simulation. Comparing cluster G_2 and G_3 in (a) and (b), the more the inflow is distributed, the lower the capacity drop is and the better the network performance. For instance, for cluster G_2 , the absolute capacity drop is equal to 0.029 veh/s (13%) for high σ in (a) and to 0.068 veh/s (30%) for low σ in (b). Last, in both cases (different values of σ), according to the cluster's identification, the network can be in a high instable intermediate state [see for instance, cluster G_2 in subplot (a) and cluster G_3 in subplot (b)], and the variability decreases when congestion is well established.

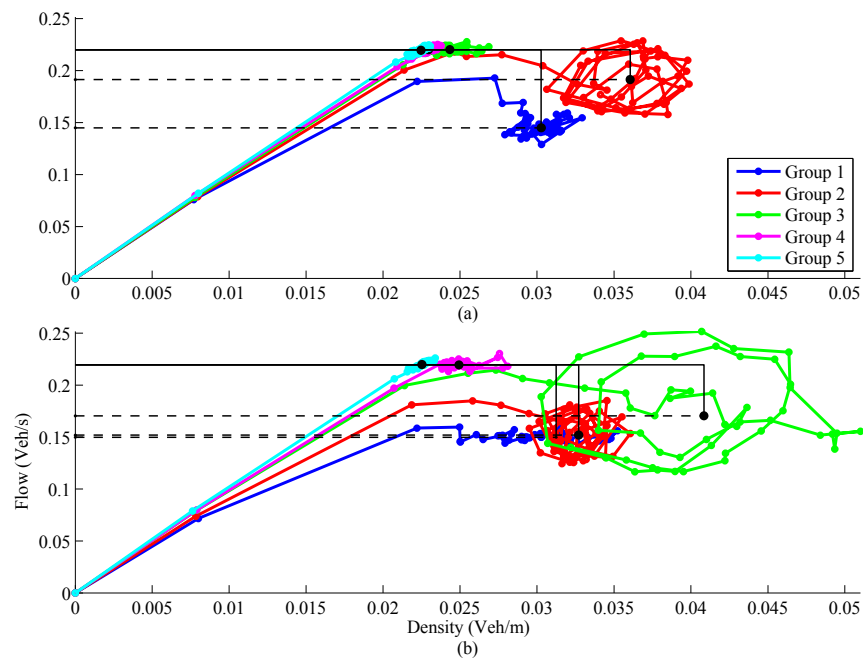


FIGURE 7 NMF D corresponding to each path's cluster for a given demand $Q_{in} = 0.15$, $\alpha = 0.15$ and for (a) a uniform path flow distribution $\sigma = 10$, and (b) a tight path flow distribution.

All previous tests have been performed attributing the same cluster's identification to the whole set of O-D pairs. In the following test, the impact of this assumption is studied. Then, Figure 8 presents the NMFD obtained for two different distributions of clusters' identification over the O-D pairs. On the upper subplot, the two border clusters are mixed while on the bottom one-mixed clusters correspond to more medium value of CF_k . First, the NMFD resulting from this cluster's distribution seems to be the mean of the two original ones: for instance mixing clusters G_1 and G_5 leads to a congested configuration but with less capacity drop. However, the combination corresponding for instance to cluster G_1 for west origins combined with cluster G_5 for north origins (green line) is not equivalent in terms of network performance to the reverse one (magenta line). Especially, the $G_1 - G_5$ mixed solution (green line) corresponds to a higher congested configuration whereas the $G_5 - 1$ one is unstable with a larger range of oscillations. It can be deduced that not only the level of overlapping plays an important role on the network level of performance but also its distribution between the O-D.

A third test is conducted to study more locally the network performance. Four distinct subnetworks as illustrated in Figure 9 upper left corner are defined, and the corresponding NMFD for each subnetwork are plotted, cluster by cluster. In this test, even if the same cluster's identification is assigned to the whole set of O-D pairs, Figure 9 shows that the behavior of the network is not spatially homogeneous depending on the level of overlapping. For instance, simulations for cluster G_1 and G_2 let appear either congested or free-flow steady state with respect to the subnetwork. Simulations for these clusters on the global network are associated to congestion, see Figure 7. Not surprisingly, subnetworks 2 and 3 appear more congested for low CF_k values than the two others quarters. This is because low CF_k values make the peripheral links more attractive. Congestion then appears close to the entries where several demand inputs aggregates.

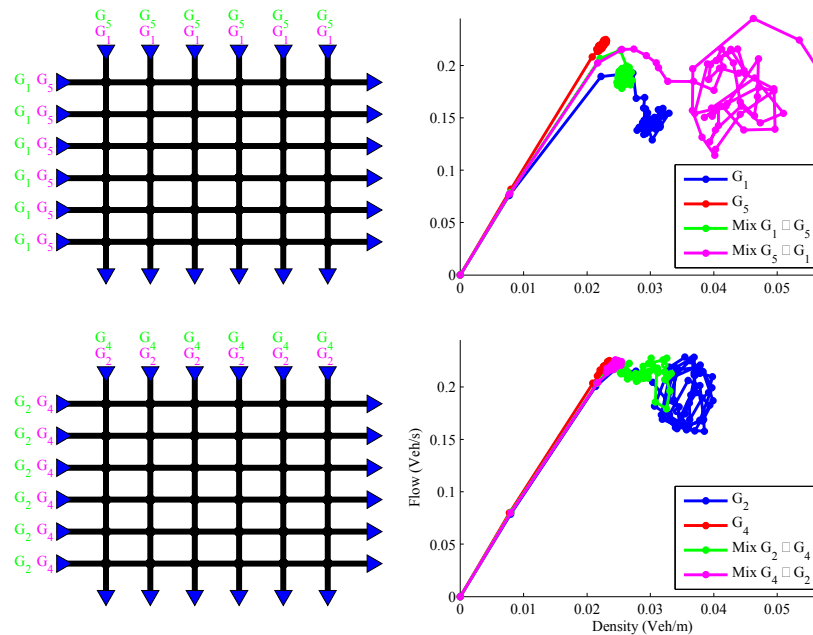


FIGURE 8 NMFD corresponding to different distributions of clusters over the O-D pairs. The demand is defined by $Q_{in} = 0.15 = \alpha$, and $\sigma = 10$ such that the inflow is uniformly distributed.

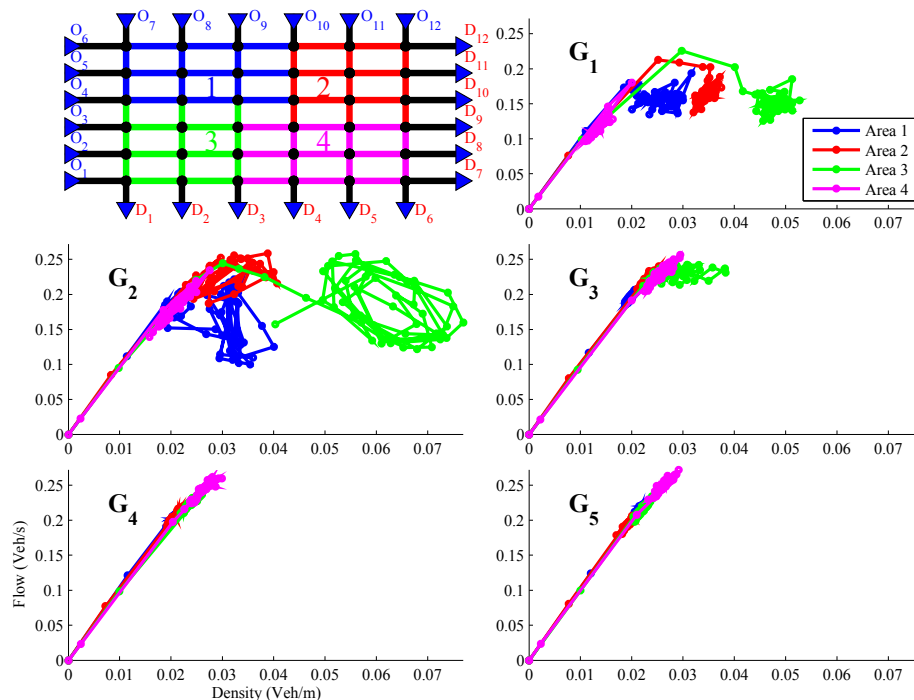


FIGURE 9 NMF D corresponding to 4 different subnetworks. The demand is defined by $Q_{in} = 0.15 = \alpha$ and $\sigma = 10$ such that the inflow is uniformly distributed.

Last, in order to study heterogeneities more in detail, Figure 10 represents the standard deviation of the density over the whole network depending on the demand for three values of σ . First, looking at the effect of the level of overlapping comparing the different lines for a given σ , it has to be noted that the less global overlapping is, the most homogeneous the network is.

Recall that low global overlapping means high CF_k values for each O-D pair in this case study (see Figure 4). Moreover, Figure 10 allows determining the range of parameters leading to the most heterogeneous zones. Indeed, it clearly appears that around the boundary of the admissible domain (in terms of constraint on the demand see Equation 1 especially for low σ value and low clusters' identification, the density on the network is highly heterogeneous with a standard deviation reaching its maximal value. This effect is dispersing when the local level of overlapping (for each O-D pair) is increasing (comparing in the first column, cluster G_1 to G_5). The last point concerns the influence of σ parameter: looking more especially at clusters G_3 to G_5 , we see that the larger σ is, the lower are heterogeneities. To conclude, the optimal solution in terms of homogeneous network seems to be when the paths are most correlated and the inflow uniformly distributed.

CONCLUSIONS

In this paper, an analysis of the effect of the local demand distribution on the network traffic conditions has been carried out, focusing on how route choices influence the level of performance of the network. In order to study route choices impact, a clustering method based on the level of paths' overlapping has been proposed to generate homogeneous sets of paths. In this

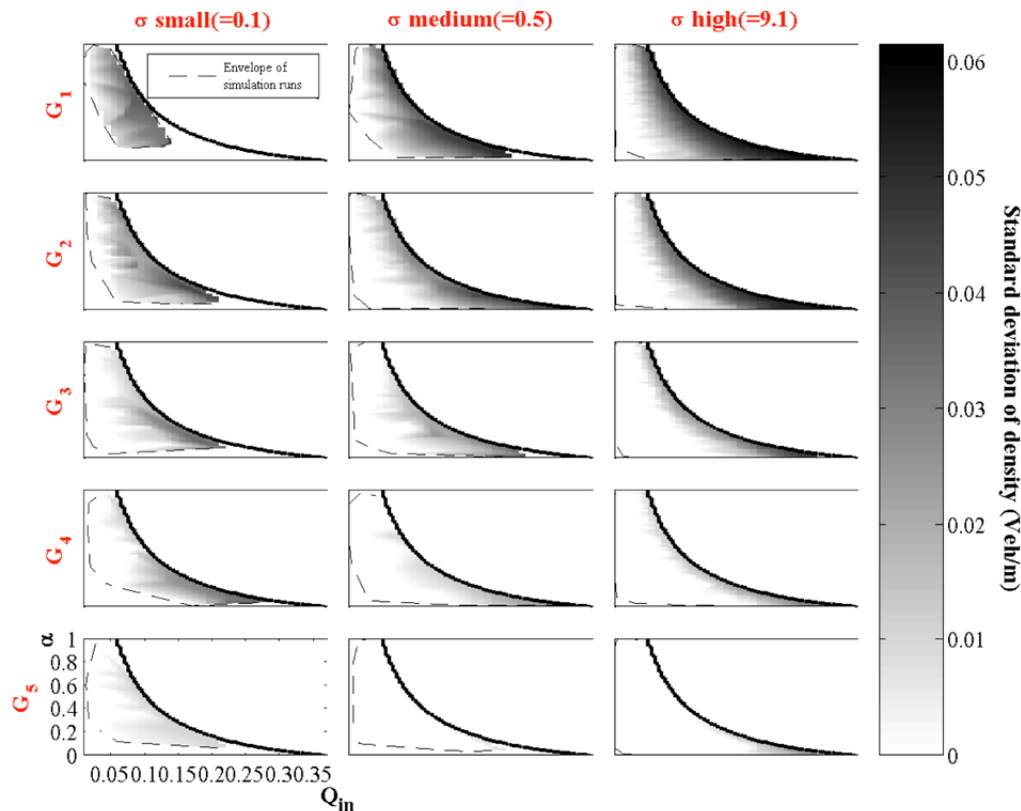


FIGURE 10 Density standard deviation depending on the demand. In columns, different values of σ and in lines, the level of overlapping from clusters G_1 to G_5 .

work a method was proposed to determine with low computational cost, the fluid/congested boundary and its sensitivity to route choice. Two levels of analysis have been performed: 1. the global network behavior through fluid–congested boundary and 2. dynamic network indicator as the NMFD and spatial distribution of densities. The different evaluated scenarios confirm that network performance is highly affected by route choice, both by paths selection and flow distribution.

An important result here is that path overlapping at the network level cannot be directly assessed through the study of paths overlapping for each O-D pair. Indeed, in this study a counterintuitive fact is observed: Low overlapping (low CF_k) for each O-D leads to the highest overlapping between paths at the network level. The question of characterizing the level of overlapping at the network level is very challenging, especially if the influence of paths overlapping and paths flow distributions are to be considered. The authors are currently investigating this question. These preliminary results and the associated framework look promising. The goal is now to improve the analysis design and perform an extensive sensitivity analysis in order to determine the most influent parameters related to route choices and demand with respect to the global network performance.

ACKNOWLEDGMENT

This work has been supported by the ADEME project CYTEDINE.

REFERENCES

1. Geroliminis, N., and C. F. Daganzo. Existence of Urban-Scale Macroscopic Fundamental Diagrams: Some Experimental Findings. *Transportation Research Part B: Methodological*, Vol. 42(9), 2008, pp. 759–770.
2. Geroliminis, N., and B. Boyaci. The Effect of Variability of Urban Systems Characteristics in the Network Capacity. *Transportation Research B*, 46(10), 2012, pp. 1576–1590.
3. Leclercq L., and N. Geroliminis. Estimating MFDs in Simple Networks with Route Choice, *Transportation Research B*, 2013.
4. Geroliminis, N., and J. Sun. Properties of a Well-Defined Macroscopic Fundamental Diagram for Urban Traffic. *Transportation Research Part B: Methodological*, Vol. 45(3), 2011, pp. 605–617.
5. Cassidy, M., K. Jang, and C. Daganzo. Macroscopic Fundamental Diagram for Freeway Networks: Theory and Observation. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2260, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 8–15.
6. Gayah, V., and C. Daganzo. Clockwise Hysteresis Loops in the Macroscopic Fundamental Diagram: An Effect of Network Instability. *Transportation Research Part B: Methodological*, Vol. 45(4), 2011, pp. 643–655.
7. Saberi M., and H. S. Mahmassani. Exploring the Properties of Networkwide Flow–Density Relations in a Freeway Network. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2315, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 153–163.
8. Knoop, V. L., S. P. Hoogendoorn, and J. W. C. Van Lint. Impact of Traffic Dynamics on the Macroscopic Fundamental Diagram. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.
9. Mahmassani, H. S., and S. Peeta. Network Performance Under System Optimal and User Equilibrium Dynamic Assignments: Implications for ATIS. In *Transportation Research Record 1408*, TRB, National Research Council, Washington, D.C., 1993, pp. 83–93.
10. Leclercq, L., and C. Becarie, C., 2011. Meso Lighthill-Whitham and Richards Model Designed for Network Applications. Presented at 91st Annual Meeting of the Transportation Research Board, Washington, D.C., 2012.
11. Lighthill, M. J., and J. B. Whitham. On Kinematic Waves II: A Theory of Traffic Flow in Long Crowded Roads. *Proceedings of the Royal Society A229*, 1955, pp. 317–345.
12. Richards, P. I. Shockwaves on the Highway. *Operations Research* 4, 1956, pp. 42–51.
13. Bovy, P. H. L., and E. Stern. *Route Choice: Wayfinding in Transport Networks*. Kluwer Academic Publishers, 1990.
14. Ben-Akiva, M., and M. Bierlaire. Discrete-Choice Methods and Their Applications to Short-Term Travel Decisions. In *Transportation Science Handbook*, 1999.
15. Szeto, W. Y., and S. C. Wong. Dynamic Traffic Assignment: Model Classifications and Recent Advances in Travel Choice Principles. *Central European Journal of Engineering*, Vol. 2(1), 2012, pp. 1–18.
16. Cascetta, E., A. Nuzzolo, F. Russo, and A. Vitetta. A Modified Logit Route Choice Model Overcoming Path Overlapping Problems. Specification and Some Calibration Results for Interurban Networks. *Proc., 13th International Symposium on Transportation and Traffic Theory*, Lyon, France, 1996, pp. 697–711.

17. Vovsha, P., and S. Bekhor. Link-Nested Logit Model of Route Choice: Overcoming Route Overlapping Problem. In *Transportation Research Record 1645*, TRB, National Research Council, Washington, D.C., 1998.
18. Bierlaire, M. The Total Demand Scale: A New Measure of Quality for Static and Dynamic Origin-Destination Trip Tables. *Transportation Research Part B: Methodological*, Vol. 36, 2002, pp. 837–850.
19. Edie, L. C. Discussion of Traffic Stream Measurements and Definitions. *Proc., 2nd International Symposium on the Theory of Traffic Flow*, OECD, Paris, France, 1963, pp. 139–154.
20. Godier, J. C., V. V. Gayah, and M. Cassidy. Inhomogeneous Flow Patterns in Undersaturated Road Networks and Implications for the MFD. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2390, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 68–75.
21. Geroliminis, N., and C. F. Daganzo. Macroscopic Modeling of Traffic in Cities. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
22. Mazloumian, A., N. Geroliminis, and D. Helbing. The Spatial Variability of Vehicle Densities as Determinant of Urban Network Capacity. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368, 2010, pp. 4627–4647.
23. Saberi, M., and H. S. Mahmassani. Empirical Characterization and Interpretation of Hysteresis and Capacity Drop Phenomena in Freeway Networks. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2391, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 44–55.
24. Tsubota, T., A. Bhaskar, E. Chung, and N. Geroliminis. Information Provision and Network Performance Represented by Macroscopic Fundamental Diagram. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.

NETWORKWIDE MODELING AND CONTROL

Macroscopic Relationship Between Networkwide Traffic Emissions and Fundamental Properties of the Network

ROOHOLAMIN SHABIIKHANI

Rutgers University

ERIC J. GONZALES

University of Massachusetts, Amherst

Emissions from traffic in networks are a growing concern, and there is a need for simple tools to estimate the relationship between network properties, traffic conditions, and the resulting aggregated emissions of pollutants such as greenhouse gases. This paper makes use of a network's macroscopic flow–density relation to approximate the following aggregated components of vehicle driving cycles: time spent cruising at free-flow speed, time spent idling, and the number of vehicle stops. The networkwide emission is estimated by multiplying these driving cycle components with associated emissions factors. The study shows that network emissions are systematically related to the network properties and vehicle density. The proposed analytical model provides an approximation of emissions within 11% of the estimates from a conventional microscopic analysis for all but the most congested traffic states. This approach allows for systematic analysis of network emissions without the need for intensive data collection and simulation.

INTRODUCTION

Road transportation is a major source of air pollutant emissions. An estimated 1.9 billion gallons (7.2 billion liters) of gasoline and \$100 billion were wasted because of fuel consumption and delays caused by traffic congestion in 2012 within the United States alone (1). In addition to wasted energy and time, urban traffic congestion contributes to networkwide emissions of air pollutants, including hydrocarbons, nitrogen oxides, carbon monoxide, and carbon dioxide. It is estimated that the on-road vehicles account for more than half of dangerous air pollutant emissions and over 30% of carbon dioxide emission in the United States (2). Reducing these emissions is important for protecting and improving human health as well as for reducing production of greenhouse gases, which are associated with global climate change. Emissions from vehicles in traffic are playing an increasingly important role in urban policy making and traffic management in large metropolitan road networks.

Most research on the relationship between traffic and pollutant emissions focuses on individual vehicles and the effect that engine technologies or driving cycles have on emissions from that vehicle. The driving cycle is the pattern of acceleration, cruising, deceleration, and idling as a vehicle traverses distance in the network. In urban environments, the design of the road network and the timing of traffic signals have systematic impacts on the driving cycles of the vehicles in the network. The traffic conditions in the network also have an impact on the

performance of vehicles, because traffic congestion causes additional stopping and idling, which directly influence the emissions from vehicles in the network. In order to evaluate, control, and reduce networkwide emissions of air pollutants, traffic emissions need to be estimated considering the nature of stop-and-go traffic in urban areas.

Although some pollutants have highly localized impacts, which require detailed models and measurements (e.g., particulate matter), greenhouse gas emissions have a global impact, and it is most important to be able to estimate the aggregated emissions from traffic in a whole network. Recent advances in modeling aggregated traffic conditions in urban networks show that a systematic relationship often exists between average vehicle flow and average vehicle density in a network. A macroscopic view of urban traffic provides a basis for making aggregated estimates of air pollutant emissions from the vehicles. This paper shows that emissions factors developed from existing microscopic emissions models can be integrated with models of aggregated traffic variables for urban networks in order to estimate the aggregated network-wide emissions of greenhouse gases from vehicles.

The paper is organized as follows. Section 2 reviews existing literature in the emission modeling and traffic flow fields. Section 3 presents the overall framework for linking aggregated traffic variables that are related to the driving cycle with emissions factors to make an aggregated emissions estimate. A detailed description of how the elements of the driving cycle are estimated from macroscopic traffic model is given in Section 4. An evaluation is presented in Section 5 by comparing the emissions estimates from the proposed model with the results from a more conventional microscopic analysis using simulation. Finally, conclusions are discussed in Section 6.

EXISTING MODELS

There are a number of existing models in the literature that focus on estimating emissions from vehicle emissions at various levels of resolution. The most detailed microscopic emissions models are based on tracking driving cycles in second by second detail, while the most aggregate models are based on broad averages in order to make regional estimates. Recent advances in traffic models that address the movements of vehicles and congestion patterns at the network level provide useful tools for analysis of aggregated traffic conditions. In the following subsections, some of the most relevant models for modeling emissions and traffic are reviewed. These form the building blocks of the proposed integrated model.

Vehicular Emissions

Existing models for vehicular emissions generally fall into two main categories: microscopic models that focus on specific movements of individual vehicles and macroscopic models that are based on aggregated data and average values. Microscopic models are the most detailed models, and they often provide instantaneous emissions estimates based on concurrent operating conditions of a an equipped vehicle or a simulation. These models typically require extensive data inputs such as second by second trajectories for each vehicle. VT-Micro (3), CMEM (4), and the project level of MOVES (5) are microscopic models that are widely used in the United States. In order to analyze the overall effect of changing a signal control system or widening a roadway, microscopic models require that a detailed microsimulation be developed to generate

the detailed trajectory of each vehicle that is then used to produce the emission estimate for each vehicle at each second. This is a time-consuming and costly process, and the data intensity and computation time make these microscopic models prohibitively burdensome for estimating emissions in large urban networks. As a result, microscopic models are typically only used in practice for analyzing small-scale projects. For greenhouse gas emissions, such detailed model outputs are not necessary in of themselves except that they tend to be more accurate than emissions estimates from macroscopic models (6).

Macroscopic emissions models are designed to estimate regional emissions from vehicles based on the average network speed, the total number of vehicles, and some assumed driving cycles (7, 8). These models require relatively few data inputs, so they are much easier to implement for large urban networks. However, these models do not account for the effect of vehicle acceleration and deceleration for stops in a way that is related to what is actually happening in the network. Macroscopic models relate average speed to a single emission rate, but in reality a single average speed could be associated with many different driving cycles ranging from a small number of long stops to a large number of short stops. These driving cycles should be associated with different emissions rates, so macroscopic models have a tendency to oversimplify the relationship between traffic patterns and emissions.

In recent years, a third type of model has emerged: mesoscopic emission models. These models do not require information about the instantaneous movements of individual vehicles, so they are not as complex and data-intensive as microscopic models. Mesoscopic models typically require aggregated traffic data that reflects the traffic conditions and congestion in the network, so they provide more accurate networkwide emission estimation in compare with macroscopic models. One example is VT-Meso, which uses link-by-link average speed, the number of vehicle stops, and the stopped delay as aggregated traffic inputs (9). The model synthesizes a typical driving cycle, and by using the microscopic VT-Micro model, it estimates the average link fuel consumption and emission rates. Overall network emissions can then be computed by aggregating the emissions on all links. Gori et al. (10) presents another mesoscopic emission model, which uses a dynamic traffic assignment model to estimate the aggregated traffic parameters, namely distance traveled at free-flow speed, the average speed of vehicles in queues, and the length of the queues. Mesoscopic models improve the accuracy of emissions estimates for larger networks, but they require inputs of aggregated traffic variables, and these need to be obtained either from a simulation or another traffic model.

Modeling Traffic in Networks

Just as emissions can be modeled at varying levels of detail, traffic models also range from microscopic models that track individual vehicle movements to macroscopic models that relate aggregated network-level variables. For the purposes of emission modeling, it is common to use microsimulation tools to construct trajectories for each vehicle that traverses an existing or hypothetical network. Although simulation models are powerful tools for investigating the complex interactions of vehicles, it is costly and challenging to build and calibrate the models appropriately (11). An alternative is to work with the classic kinematic wave model (12, 13) that makes some simplifying assumptions about the variability of driver and road characteristics but can describe the evolution of traffic states on a road segment by tracking the interfaces between traffic states over space and time. The benefit of this analytical approach is that a wide variety of traffic scenarios can be evaluated in a robust and consistent way with far less data and

computational complexity than a microsimulation. At the level of intersections and individual arterials, kinematic wave theory has been a basis of traffic modeling for decades.

For networks that are homogeneous and well connected and on which demand is uniformly spread, a consistent relationship between average network flow and average network density has been shown to exist in theory (14, 15), in simulations (16), and in the real world (17, 18). This relation is often referred to as the macroscopic fundamental diagram (MFD) or network-level fundamental diagram. The size and shape of the MFD depends primarily on the physical properties of the network including the saturation flow rate, block length, and traffic signal settings (e.g., cycle length, duration of signal phases, and signal offsets). This aggregate relation of traffic variables is useful for a network manager, because it can be used to monitor the network performance or implement control strategies to increase throughput and decrease delays in the system (19). An additional objective may be to reduce aggregated fuel consumption and emissions in a network, but this application of networkwide traffic models has received less attention in the literature.

Since the critical input for emissions models is an accurate driving cycle, traffic models need to relate the time that vehicles spend accelerating, cruising, decelerating, and idling to the traffic conditions on the roadway. An arterial-level model has been developed to estimate emissions assuming that some traffic data, such as flows and number of vehicle stops, are measured directly from links in the network and then estimating the other relevant parts of the driving cycle (20). Another recent model uses kinematic wave theory to make analytical estimates of the entire driving cycle for traffic on a single link approaching an isolated intersection (21). The model proposed in this paper is intended to go a step further to estimate emissions based on aggregated traffic characteristics using the MFD and physical characteristics of the network.

INTEGRATED TRAFFIC EMISSION MODEL FOR A NETWORK

The proposed modeling framework builds on the Integrated Traffic Emission Model (ITEM) presented in Shabihkhani and Gonzales (21), which connects an analytical model of traffic approaching an isolated intersection with emission factors from a microscopic emission model. That study shows that reliable predictions of emissions at a signalized intersection can be made using kinematic wave theory to estimate the amount of time vehicles spend idling, the time spent cruising, and the number of times that vehicles stop per vehicle distance traveled. The model of networkwide emissions presented in this paper is structured with the same two components: a traffic model to estimate aggregated traffic parameters and a set of emissions factors to convert the driving cycle into an emissions estimate.

The trajectories of vehicles approaching an intersection or traversing a network have repeating patterns of cruising at the free-flow speed, v_f , idling while stopped, and decelerating and then accelerating between speeds v_f and 0 for every stop. Therefore, three components of the driving cycle that must be estimated from the traffic model in order to account for emissions from the vehicles: the time spent cruising per distance traveled, T_c ; the time spent idling per distance traveled, T_i ; and the number of times that vehicles must stop per distance traveled, n . The total emissions per vehicle distance traveled, E , is then calculated by multiplying these components by the appropriate emissions factors:

$$E = e_c T_c + e_i T_i + e_s n \quad (1)$$

where e_c is the emission of interest per unit cruising time, e_i is the emission of interest per unit idling time, and e_s is the total emission of interest associated with a complete deceleration from v_f to 0 and a complete acceleration from 0 to v_f .

In order to make accurate emissions estimates, it is important to have accurate estimates of the components of the driving cycle (T_c , T_i , and n) and accurate emission factors (e_c , e_i , and e_s). The analysis in this paper will focus on investigating simple homogeneous networks in which the MFD is known to be applicable so that we can focus on using the MFD to estimate driving cycles. Then these driving cycles will be used to estimate emissions. The details about how to estimate the driving cycle from the macroscopic traffic data are presented in Section 4. How to obtain appropriate emissions factors, which are important for analysis of isolated intersections or larger networks, will also be considered. Although the method may be applied to measured or simulated vehicle data from any road or network, this investigation will use a simulation approach to study the performance of idealized networks.

Traffic Simulation

The first step to estimating emissions factors with a microscopic emissions model is to obtain high-resolution vehicle trajectories that show speed and acceleration at a fine temporal resolution (e.g., every second). In the field, trajectories can be measured from equipped vehicles, but a simulation model is useful for considering a wider range of traffic conditions, many of which may not be part of a measured data set. In order to represent the ideal homogeneous network conditions under which a consistent MFD has been proven to exist, a simple ring network has been constructed using Aimsun that is consistent with the theoretical assumptions in Daganzo (14) and Daganzo and Geroliminis (15). The ring with a single intersection is representative of a long arterial or network with homogeneous traffic conditions and traffic signals with no offset.

In the ring model, a constant number of vehicles in the system correspond to a constant density. The full range of possible densities from an empty network up to a complete jam are systematically analyzed by loading the ring with a specific number of vehicles and then running the simulation to measure aggregated network flow and extract vehicle trajectories. Feeding each trajectory into a microscopic emission model provides a second-by-second estimate of the emissions from each vehicle. Aggregating the emissions from all the individual vehicle trajectories provides an estimate of the networkwide emissions following the conventional detailed microscopic approach. In this paper, the project level of MOVES (5) is used as the microscopic emission model, but the same method could be used with any microscopic model that uses vehicle trajectories as the model input.

Estimation of Emission Factors

The goal is to estimate emission factors for each component of the driving cycle, so a sample of trajectories is parsed into cruising, idling, acceleration, and deceleration. This process requires that thresholds be defined to distinguish between slight oscillations in speed and larger changes that are associated with accelerations and decelerations associated with stopping. The following criteria were used to parse the trajectories in Shabihkhani and Gonzales (21), and they are used again in this study:

1. A vehicle is considered to be stopped and idling whenever the speed is slower than 1 mph (1.6 km/h).
2. A vehicle is considered to be accelerating or decelerating when the following conditions hold: the absolute value of the rate of acceleration exceeds 0.2 mph/s (0.3 km/h/s); the speed changes by at least 5 mph (8 km/h); the duration of the acceleration or deceleration lasts at least 2 s; an intermediate period of opposite acceleration does not exceed 1 s; and an intermediate period of low acceleration does not exceed 3 s.
3. The remaining time, the vehicle is moving at steady enough speed that it is considered to be cruising.

These criteria were identified because they provided the closest match between the number of stops counted with the automated procedure and the number of manually counted stops from empirically measured and simulated trajectories. Although this parsing process may appear complicated, the important thing is collect observations of enough vehicle trajectories to obtain a good estimate of the average idling, cruising, accelerating, and decelerating behaviors.

Once the trajectories have been broken into each of the components of the driving cycle, each trajectory segment has a duration and is analyzed with a microscopic emission model to estimate the corresponding vehicle emission. For the idling and cruising, the results are simply averaged to obtain an average emission rate for each second of idling and each second of cruising. For the accelerations and decelerations the duration and total emission are both important quantities. Each stop requires that a vehicle decelerate and accelerate, so the sum of the deceleration and acceleration durations are the period of time when vehicles are neither cruising nor idling. The cycle of decelerating and accelerating for a stop is associated with a quantity of pollutants emitted per vehicle stop.

This paper evaluates the proposed analytical model with a number of different network scenarios in which the free-flow speed is $v_f = 53$ km/h. The project level of MOVES was used to analyze a sample of trajectories extracted from an Aimsun simulation of a ring-shaped network as described in Section 3.1. The emissions of interest for our study are greenhouse gases, because these are global pollutants that are most important to estimate in aggregate for a network. The relevant unit of measure for greenhouse gases is grams of carbon dioxide equivalents (gCO₂eq) because, this represents the global warming of all greenhouse gases emitted from the vehicles in terms of an equivalent amount of CO₂. The emissions factors for this case are $e_c = 2.187$ gCO₂eq/s, $e_i = 0.881$ gCO₂eq/s, $e_s = 48.876$ gCO₂eq/stop, and the average duration of a deceleration and acceleration cycle is $\tau = 22$ s.

ANALYTICAL MODEL FOR NETWORKWIDE TRAFFIC VARIABLES

Existing macroscopic models for networkwide traffic conditions relate the average network flow, q , to the average network density, k . These two variables imply the average speed of vehicles in the network, v , by the well-known relation:

$$v = q/k \quad (2)$$

These variables alone provide a lot of useful traffic information about the capacity of a network and the delays that drivers in the network experience. Ongoing research is being

conducted to better understand the behavior of the macroscopic flow–density relation for different types of realistic networks. For the proposed model it is supposed that the MFD for a network is known or has been measured use it to provide an analytical approximation for the idling time, cruising time, and number of stops for vehicles in the network. The goal is to develop a model with sufficient detail to estimate aggregated emissions in the network without the need to track the details of each vehicle’s movements.

As presented in Section 3, the complexities of a second-by-second vehicle trajectory can be simplified into three key parts of the driving cycle that are related to emissions: time spent moving at the cruising speed per vehicle-distance, T_c ; time spent idling per vehicle-distance, T_i ; and the number of times that vehicle stops per vehicle-distance, n . First considered is how T_c and T_i can be estimated if n is known. Then it will be considered how the number of stops per distance can be estimated as well.

Suppose that traffic on a homogeneous network has a triangular fundamental diagram with free-flow speed of v_f . If it is ignored for the moment the range of speeds that are associated with acceleration and deceleration, vehicles will have piecewise linear trajectories with speed v_f while moving (i.e., cruising) or stopped while idling. All travel time for vehicles can be classified as effectively cruising or effectively idling. The kinematic waves associated with these idealized trajectories are the same as the aggregated dynamics of traffic with more realistic acceleration and deceleration patterns (12, 13).

Every vehicle that stops must decelerate from v_f to 0 and then accelerate from 0 back to v_f . The duration of the deceleration is τ_d and the duration of the acceleration is τ_a , and these values depend on the behavior of drivers in a particular network. If the deceleration and acceleration are at constant rates, then half of τ_d and τ_a is effectively cruising time and the other half is effectively idling time. Figure 1 shows how a piecewise linear trajectory and a more realistic trajectory with constant rates of deceleration and acceleration. For simplicity, we will

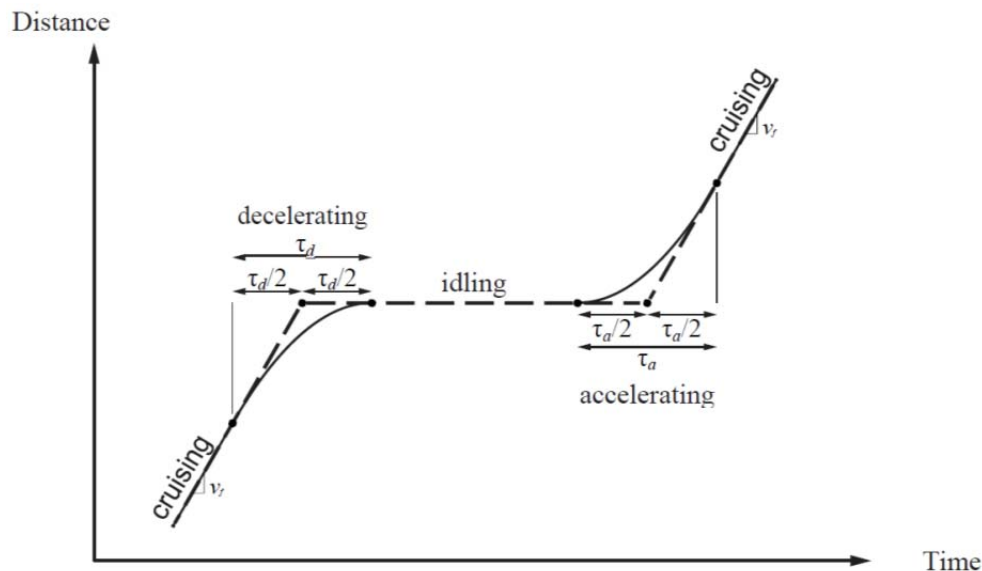


FIGURE 1 Relationship between a trajectory with constant deceleration and acceleration rates (solid) and a piecewise linear trajectory simplified to effective cruising and effective idling (dashed).

consider a single time associated with the cycle of deceleration and acceleration for each vehicle stop $\tau = \tau_d + \tau_a$. Therefore the each stop reduces the actual time spent cruising by $\tau/2$ and the actual time spent idling by $\tau/2$. It is important to account for τ when modeling traffic emissions, because the emission rates for cruising and idling should be multiplied by the actual cruising and idling times rather than the effective times.

The effective cruising time per unit distance is simply the inverse of the free-flow cruising speed, because no distance is traversed while idling. The actual cruising time per unit distance is then calculated by reducing the effective cruising time by half of deceleration and acceleration time for each stop:

$$T_c = \frac{1}{v_f} - \frac{\tau}{2} n \quad (3)$$

where n is the number of times a vehicle stops per unit distance traveled.

The effective idling time is the difference between the total travel time per unit distance, which is the inverse of the average traffic speed, and the effective cruising time. The actual idling time per unit distance is again calculated by reducing the effective idling time by the other half of the deceleration and acceleration time per stop:

$$T_i = \frac{1}{v} - \frac{1}{v_f} - \frac{\tau}{2} n \quad (4)$$

In many cases, it may be possible to measure n from the same data source used to obtain the estimated macroscopic traffic state k and q (i.e., traffic data from probe vehicles could provide an indication of this value). In the absence of direct measurements, it is useful to be able to express the number of stops analytically. Although an individual vehicle makes a discrete number of stops per distance traveled, this could vary across vehicles or road segments. Therefore, it is useful to be able to have an analytical approximation for n .

The simplest approximation is simply to suppose that on average vehicles are stopped once per cycle. The average distance traveled during a signal cycle of length C is vC , so the number of stops per distance is given by

$$n = \frac{1}{vC} \quad (5)$$

This approximation is appropriate when the signal offset is 0, and especially when the duration of the red signal exceeds the time required to travel the length of a block at free-flow speed: $C - G \geq \ell/v_f$, where ℓ is the length of a block. When the red phase is sufficiently long, a vehicle will always have to stop once per cycle when caught at a red signal. When block lengths or signal times are short enough that this inequality is violated, it is possible for some vehicles to traverse the network without stopping during every cycle, which is a possible source of errors.

Since the MFD is a property of a specific network, the flow can be expressed as a function of density, $q = Q(k)$. The shape and size of $Q(k)$ depends on the network characteristics (e.g., saturation flow, jam density, and block length) and traffic signal settings (e.g., cycle length and green ratio). Therefore the average speed of vehicles in the network can be expressed as a function of density, so (2) becomes $v(k) = Q(k)/k$. The emissions in a network are estimated by evaluating T_c , T_i , and n with $v(k)$ and substituting the resulting driving cycle components into Equation 1.

EVALUATION WITH AN IDEALIZED NETWORK

The proposed analytical model is evaluated by constructing an idealized network in a microsimulation and then comparing the emissions estimates with the results of a conventional microscopic emissions analysis. The accuracy of analytical approximations for the MFD itself is beyond the scope of this paper, so the analytical approximations are made assuming that the MFD is measured and known. A ring model is used to represent an idealized homogeneous network as explained in Section 3.1. It is from this simulation that the empirical MFD is measured, and the detailed vehicle trajectories are also extracted in order to calculate the modeling error relative to the conventional microscopic modeling approach.

First a comparison between the analytical modeling approach and the conventional simulation approach is investigated for a base case network. Then, the effect of varying one network parameter, the green ratio, is demonstrated using the analytical model to show how the proposed modeling approach can be used to evaluate changes to the system. Finally, an error analysis is conducted to compare the performance of the analytical model relative to the conventional microscopic approach shows that for a wide variety of network characteristics and traffic states.

COMPARISON OF ANALYTICAL AND SIMULATION MODEL RESULTS

The base case network that is used to illustrate the performance of the proposed analytical model has the following properties: free-flow speed, $v_f = 53$ km/h; saturation flow, $s = 1,900$ veh/lane-h; jam density, $k_j = 200$ veh/lane-km; green ratio (length of green phase divided by signal cycle length), $G/C = 0.50$; signal cycle length, $C = 60$ s; block length, $\ell = 0.30$ km; and no signal offset. Running the simulation for a range of densities between 0 and k_j , the average network flow q is plotted for each density k in Figure 2. The points in the figure indicate the measurements from the simulation, and we will suppose that $Q(k)$ is the empirical curve connecting these points (shown as the solid line).

Using average network speed at each density, $v(k) = Q(k)/k$, the number of stops is estimated using Equation 5. Figure 3a shows the analytically estimated value of n (solid line) and the number of stops determined by analysis of the simulated vehicle trajectories (dots) as described in Section 3.2. The plot shows that the analytically estimated number of stops has a similar and close trend to simulated values, especially at low densities ($k < 75$ veh/lane-km) associated with the free-flow branch of the MFD. At greater densities the number of stops observed in simulation start to grow faster than the analytical prediction, because the interactions between vehicles as conditions become congested create some additional stop-and-go waves that are not accounted for in the simple model. At the highest densities ($k > 175$ veh/lane-km), where traffic is nearly completely jammed, the estimated number of stops per distance soars while the observed number of stops actually declines. This is a result of the fact that in extremely congested conditions, vehicles move so little during each cycle that the trajectories do not trigger the necessary thresholds for the stops to get counted.

The analytically computed values for n are then used along with the values of $v(k)$ to estimate the time per distance spent cruising, based on Equation 3, and idling, based on Equation 4. Figure 3b shows the analytically estimated idling time (solid line) and the idling time measured from the simulated trajectories (dots). The analytical approximation fits closely with the simulated values.

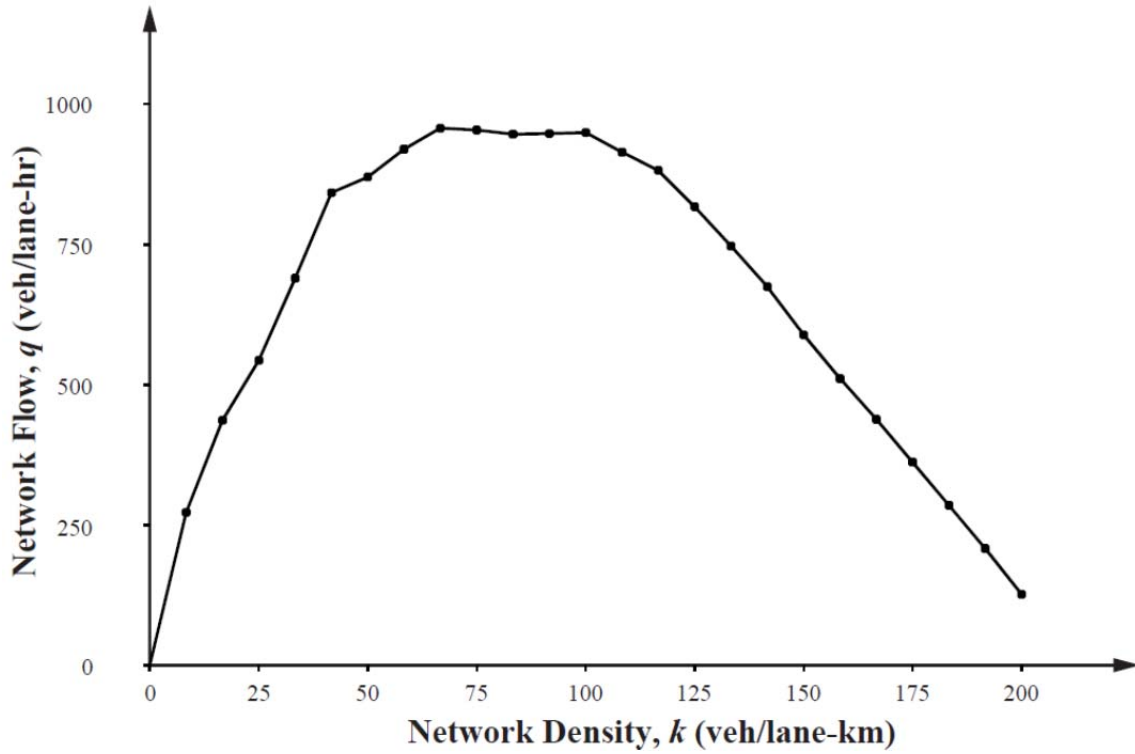


FIGURE 2 Network flow–density relation (MFD) measured from the simulation of an idealized ring network ($G/C = 0.50$; $C = 60$ s; $\ell = 0.30$ km).

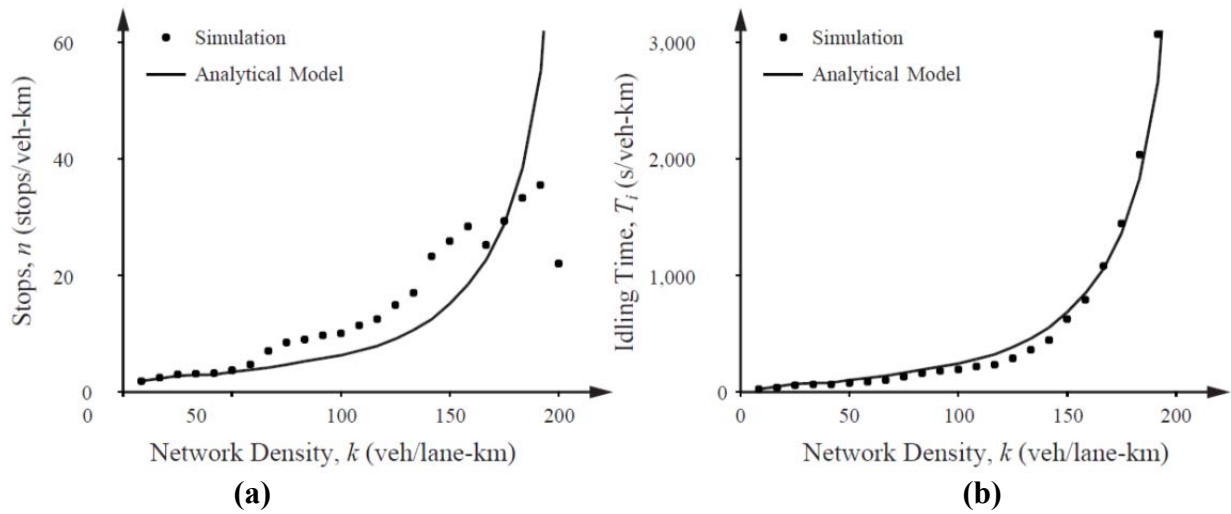


FIGURE 3 Components of the driving cycle estimated using the analytical model and measured from simulation (a) number of stops and (b) idling time ($G/C = 0.50$; $C = 60$ s; $\ell = 0.30$ km).

Since the idling time and cruising time are calculated by subtracting the duration of the deceleration and acceleration cycles associated with each stop, errors in the estimated number of stops contribute to errors in the estimated values of T_c and T_i . The values of k where stops are underestimated also have overestimated values of T_i and vice versa. The error that affects the idling time (as shown Figure 3) also affects the cruising time estimates (not shown) in a similar way.

The total greenhouse gas emission per vehicle distance traveled is calculated by multiplying each of the estimated driving cycle components by the associated emission factors as show in Equation 1. These results can be compared with the outcome of a conventional microscopic emissions analysis using the simulated vehicle trajectories. A comparison of the analytically estimated emissions (solid line) and the aggregated simulation output (dots) is shown in Figure 4. The close agreement between the analytical macroscopic model and the detailed simulation model occurs because aggregating the emissions from all vehicle trajectories together has the effect of averaging out variations from vehicle to vehicle.

Variation of Signal Timing

The proposed analytical model is particularly useful for comparing the performance of networks with different characteristics. One example is to consider the effect that changing signal timings has on the emissions from traffic in a network. Using all the same network parameters as the base case presented in Section 5.1, an evaluation of the effect of changing the green ratio is conducted by changing only the value of G/C . Figure 5a shows the MFD for each of the green ratios $G/C \in \{0.25, 0.50, 0.75\}$. The middle value is the same base case presented in Figure 2.

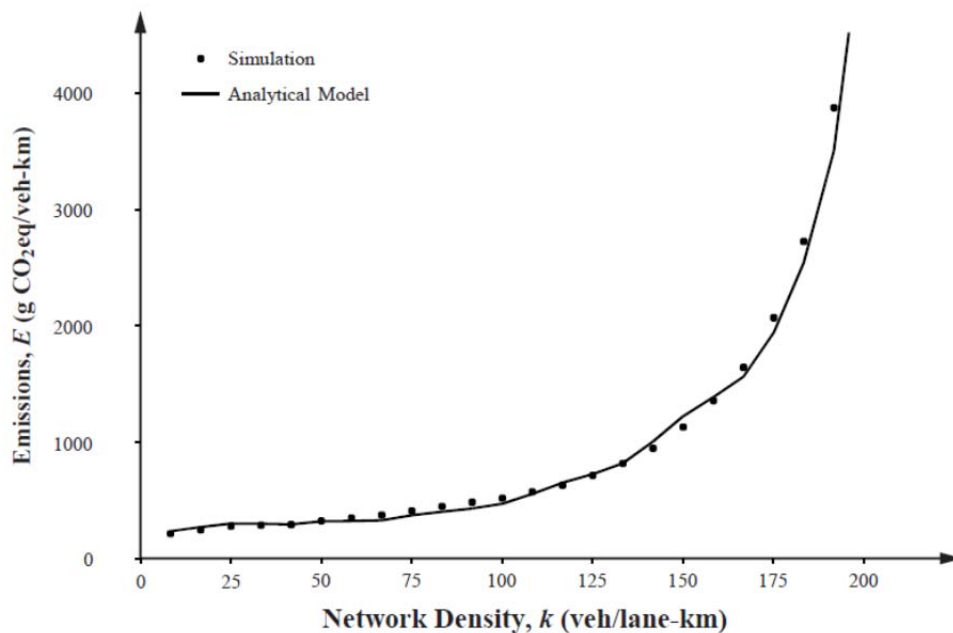


FIGURE 4 Networkwide emissions estimated using an analytical model based on the MFD and estimated using detailed trajectories from a simulation and microscopic emission analysis ($G/C = 0.50$; $C = 60$ s; $\ell = 0.30$ km).

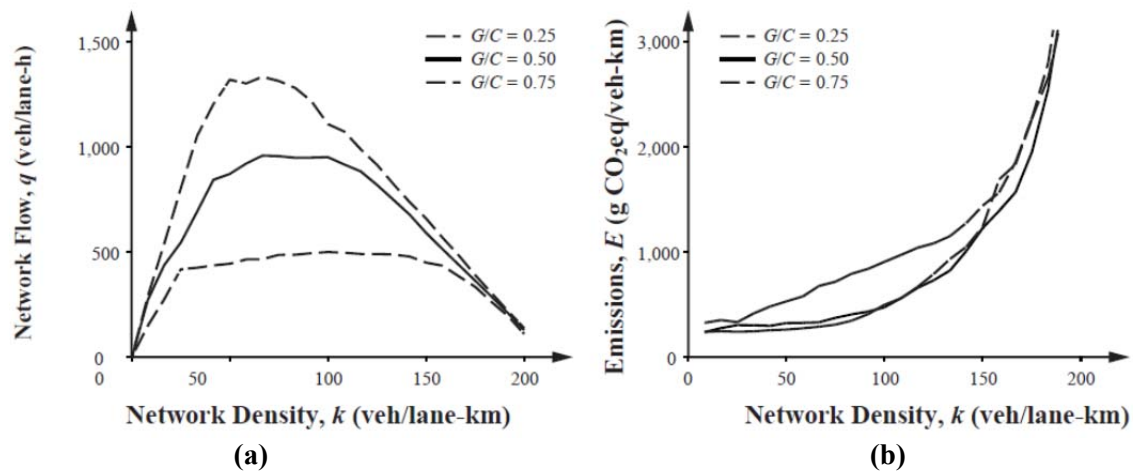


FIGURE 5 Comparison of MFD and analytically estimated emissions for varying green ratios, $G/C \in \{0.25, 0.50, 0.75\}$: (a) network flow (MFD) and (b) network emissions.

The effect of G/C on the MFD is not surprising, because a longer green phase within the cycle allows a greater flow of vehicles to traverse the network. The most restrictive green time ($G/C = 0.25$) is associated with a low network capacity, and a constant flow that is associated with a wide range of densities. The analytically estimated emissions for each of the cases are shown in Figure 5b. The results show that the more restricted green ratio is associated with greater emissions per vehicle distance traveled, but there is not a big difference between $G/C = 0.50$ and $G/C = 0.75$.

The ability to compare scenarios based only the MFD is useful because detailed trajectories do not need to be extracted and analyzed with the microscopic emission model for each case considered. A similar method can be applied to changing other network parameters such as the cycle length, C , and block length, ℓ . All of these cases are associated with the same free-flow speed, v_f , so the set of the emission factors (e_c , e_i , and e_s) and the duration of each acceleration and deceleration cycle (τ) remains the same as the base case. If the free-flow speed in the network were to change, these factors would have to be re-estimated.

Model Errors

In order to assess the accuracy and robustness of the proposed analytical model, an error analysis has been performed to compare the estimated emissions from the analytical model with the results of a detailed simulation and microscopic emission analysis. The accuracy is quantified by calculating the percent error of each analytically calculated emission value relative to the simulated result.

Starting from the base case presented in Section 5.1 with $G/C = 0.50$, $C = 60$ s, and $\ell = 0.30$ km, a systematic error analysis was conducted for each of the following variations in isolation: the green ratio, $G/C \in \{0.25, 0.50, 0.75\}$; the signal cycle length, $C \in \{30, 60, 120\}$ s; and the block length, $\ell \in \{0.15, 0.30, 0.60\}$ km. For each case a separate ring simulation was constructed to generate the MFD for the analytical approximation and to generate the detailed vehicle trajectories for the conventional microscopic analysis. The percent error of the proposed analytical approach relative to the conventional microscopic simulation approach is summarized in Table 1.

TABLE 1 Percent Error of Emissions Estimate from the Aggregated Analytical Emission Model Relative to the Microscopic Simulation Model (Base Case: $G/C = 0.5$; $C = 60$ s; $\ell = 0.30$ km)

Network Properties			Network Density, k (veh/lane-km)				
G/C	C (s)	ℓ (km)	25	50	100	150	200
Variation of the Green Ratio							
0.25	60	0.30	2.1%	4.9%	0.9%	1.7%	15.9%
0.50	60	0.30	-7.7%	10.5%	10.5%	-7.6%	19.5%
0.75	60	0.30	-8.4%	-7.2%	-7.2%	-17.4%	22.4%
Variation of the Signal Cycle Length							
0.50	30	0.30	9.3%	10.1%	5.5%	0.4%	49.7%
0.50	60	0.30	-7.7%	1.3%	10.5%	-7.6%	19.5%
0.50	120	0.30	-11.0%	-10.0%	0.2%	-1.8%	1.1%
Variation of the Block Length							
0.50	120	0.15	-5.0%	8.5%	6.7%	-1.2%	22.6%
0.50	120	0.30	-7.7%	1.3%	10.5%	-7.6%	19.5%
0.50	120	0.60	-10.3%	-7.5%	1.4%	0.2%	22.4%

The network scenarios are clustered into three groups, each group showing the results of varying one of the network variables. The center row of each cluster is the base case so that the effect on the percent error from increasing and decreasing each variable can be compared one at a time. In almost all cases when the network is not completely jammed ($k < 200$ veh/lane-km), the model is within 11% of the simulated value. These errors do not appear to have a systematic bias and the magnitudes are small relative to the variation in emission rates for different values of k as shown in Figures 4 and 5b. Therefore, the proposed analytical model provides a good approximation for the detailed microscopic estimates.

Only at the jam density ($k = 200$ veh/lane-km) are the errors very large and consistently positive. These large errors occur when the network is near a state of complete gridlock, because the model predicts a large number of stops but the traffic moves so little with each cycle that the vehicle trajectories in the simulation never move faster than a slow crawl. Fortunately, these extremely jammed conditions are rare, and the model performs well for a wide range of congested traffic conditions and a wide range of network characteristics.

CONCLUSION

A model has been proposed that makes use of the macroscopic relationship between average flow and density known as the MFD to make analytical estimates of the network-wide emissions from traffic. A robust relationship is shown between the components of that driving cycle that are associated with vehicular emissions and the fundamental properties of the network. Aggregated traffic parameters are used to identify a typical driving cycle. The components of the driving cycle per vehicle distance traveled (i.e., cruising time, idling time, and number of stops) are estimated based on the aggregated flow–density relation (MFD), the free-flow speed in the network, the

duration of a typical acceleration and deceleration associated with a vehicle stop, and the signal cycle length. These components are then multiplied by emission factors that are developed using a detailed microscopic emission model, such as the project level of MOVES.

The ITEM that has been presented and evaluated in this paper links macroscopic traffic flow models with microscopic emissions models in order to exploit the strengths of each modeling approach. Conventional microscopic traffic emissions modeling requires detailed data for individual trajectories, which must either be measured in the field or generated with a microsimulation, in order to make detailed emissions estimates. This is not practical for estimating emissions in large urban networks, but that macroscopic emissions models that are currently available do not adequately account for the effect that properties of the road network have on driving cycles and the resulting emissions estimates. The proposed modeling approach addresses this challenge by making use of state-of-the-art macroscopic traffic models that are sensitive to properties of the network such as the lane capacities, block lengths, and traffic signal timings. By making use of the MFD, which embodies the effects of network properties on the aggregated flow–density relation, networkwide emissions can be reliably estimated for a wide range of traffic conditions without the need for extensive simulations and trajectory analysis.

The effect of network characteristics and traffic dynamics on real MFDs is currently a topic of extensive research. The flow–density is known to exist and be robust for idealized homogeneous networks, so this was used to demonstrate the potential for using a macroscopic approach to approximate driving cycles in the network. The shape of the MFD has been studied for various types of networks and have been developed (15, 16, 22), but it is supposed that this relation is either measured or determined by some other method. Given the traffic state on the MFD, a few other network characteristics (v_f , C , and τ), and the emission factors (e_c , e_i , and e_s), the ITEM has been shown to approximate the vehicular emissions within 11% of the values from microscopic analysis of simulated trajectories for all but the most congested traffic states.

This proposed model is useful for monitoring emissions in real networks because traffic states can be monitored using data collected from many different sources, including vehicle probes, mobile phones, and fixed detectors. The same data that is useful for monitoring traffic and implementing efficient traffic control system can also be used to estimate networkwide emissions without simulations or extensive additional data collection. Furthermore, the analytical approach provides a tool for systematically analyzing the effect of changes to the network on emissions by tracking the effect on the MFD. While this paper has focused on demonstrating the potential of this integrated model with an idealized ring-shaped network, additional work is needed to determine how well the modeling approach applies to more realistic networks that may have turning vehicles, signal offsets, or inhomogeneous signal timings and block lengths. Nevertheless, the proposed model has value because it provides a less data-intensive way to estimate aggregated network emissions, which is especially important for tracking pollutants like greenhouse gases that have a global impact.

ACKNOWLEDGMENT

This work is supported by the Rutgers University Center for Advanced Infrastructure and Transportation, a Tier 1 consortium under a grant from the U.S. Department of Transportation Research and Innovative Technology Administration.

REFERENCES

1. *A New Economic Analysis of Infrastructure Investment*. U.S. Department of the Treasury, 2012, <http://www.treasury.gov/resource-center/economic-policy/Documents/20120323InfrastructureReport.pdf>.
2. *Light-Duty Automotive Technology, Carbon Dioxide Emissions, and Fuel Economy Trends: 1975 Through 2013*. Technical report. U.S. Environmental Protection Agency, 2013.
3. Rakha, H., M. Van Aerde, K. Ahn, and A. Trani. Requirements for Evaluating Traffic Signal Control Impacts on Energy and Emissions Based on Instantaneous Speed and Acceleration Measurements. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1738*, TRB, National Research Council, Washington, D.C., 2000, pp. 56–67.
4. Barth, M., F. An, T. Younglove, C. Levine, G. Scora, M. Ross, and T. Wenzel. *Final Report: Development of a Comprehensive Modal-Emissions Model*. NCHRP 25-11, National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., 2000.
5. *Motor Vehicle Emission Simulator (MOVES): User Guide for MOVES2010b*. Technical report. U.S. Environmental Protection Agency, 2010.
6. Rakha, H., K. Ahn, and A. Trani. Comparison of MOBILE5a, MOBILE6, VT-MICRO, and CMEM Models for Estimating Hot-Stabilized Light-Duty Gasoline Vehicle Emissions. *Canadian Journal of Civil Engineering*, 30(6), 2003, pp. 1010–1021.
7. Akcelik, R. An Interpretation of the Parameters in the Simple Average Travel Speed Model of Fuel Consumption. *Australian Road Research*, 15(1), 1985.
8. Bai, S., D. Eisinger, and D. Niemeier. MOVES vs. EMFAC: A Comparison of Greenhouse Gas Emissions Using Los Angeles County. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
9. Yue, H. *Mesosopic Fuel Consumption and Emission Modeling*. PhD dissertation. Virginia Polytechnic Institute and State University, Blacksburg, 2008.
10. Gori, S., S. La Spada, L. Mannini, and M. Nigro. Within-Day Dynamic Estimation of Pollutant Emissions: A Procedure for Wide Urban Network. *Procedia—Social and Behavioral Sciences*, Vol. 54, 2012, pp. 312–322.
11. Dowling, R., A. Skabardonis, J. Halkias, G. McHale, and G. Zammit. Guidelines for Calibration of Microsimulation Models: Framework and Applications. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1876*, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 1–9.
12. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 229(1178), 1955, pp. 317–345.
13. Richards, P. I. Shock Waves on the Highway. *Operations Research*, 4(1), 1956, pp. 42–51.
14. Daganzo, C. F. Urban Gridlock: Macroscopic Modeling and Mitigation Approaches. *Transportation Research Part B*, 41(1), 2007, pp. 49–62.
15. Daganzo, C. F., and N. Geroliminis. An Analytical Approximation for the Macroscopic Fundamental Diagram of Urban Traffic. *Transportation Research Part B*, 42(9), 2008, pp. 771–781.
16. Ji, Y., W. Daamen, S. Hoogendoorn, S. Hoogendoorn-Lanser, and X. Qian. Investigating the Shape of the Macroscopic Fundamental Diagram Using Simulation Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2161*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 40–48.
17. Geroliminis, N., and C. F. Daganzo. Existence of Urban-Scale Macroscopic Fundamental Diagrams: Some Experimental Findings. *Transportation Research Part B*, 42(9), 2008, pp. 759–770.
18. Buisson, C., and C. Ladier. Exploring the Impact of Homogeneity of Traffic Measurements on the Existence of Macroscopic Fundamental Diagrams. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2124*, Transportation Research Board of the National Academies, 2009, pp. 127–136.

19. Geroliminis, N., J. Haddad, and M. Ramezani. Optimal Perimeter Control for Two Urban Regions with Macroscopic Fundamental Diagrams: A Model Predictive Approach. *IEEE Transactions on Intelligent Transportation Systems*, 14(1), 2013, pp. 348–359.
20. Skabardonis, A., N. Geroliminis, and E. Christofa. Prediction of Vehicle Activity for Emissions Estimation Under Oversaturated Conditions Along Signalized Arterials. *Journal of Intelligent Transportation Systems*, 17(3), 2013, pp. 191–199.
21. Shabihkhani, R., and E. J. Gonzales. Analytical Model for Vehicle Emissions at a Signalized Intersection: Integrating Traffic and Microscopic Emissions Models. Presented at 92nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2013.
22. Gayah, V. V., and C. F. Daganzo. Clockwise Hysteresis Loops in the Macroscopic Fundamental Diagram: An Effect of Network Instability. *Transportation Research Part B: Methodological*, 45(4), 2011, pp. 643–655.

NETWORKWIDE MODELING AND CONTROL

Existence, Stability, and Mitigation of Gridlock in Beltway Networks

WEN-LONG JIN

University of California, Irvine

Previous studies have shown that the gridlock state can arise in a beltway network. However, no closed-form formulations were provided for traffic dynamics in a beltway network, and the existence and stability of gridlock were not established systematically. This paper first presents the network kinematic wave model for traffic dynamics in a rotationally symmetric beltway network, including the merging and diverging models. Then it demonstrates that gridlock is always a stationary solution with fixed exiting ratios. Further, it shows that the gridlock state may be stable or unstable, depending on the merging priority and the exiting ratio. Additionally, the mechanisms and limitations of existing mitigation strategies are discussed and a new adaptive driving strategy is proposed. Finally, the paper concludes with future research directions.

INTRODUCTION

A beltway network is a ring road with multiple entrance and exit links, as shown in Figure 1, that can be embedded in both freeway and arterial networks. A gridlock state in a beltway network is defined as a traffic state when the total density is positive, but the flow-rate is zero. A trivial case is when all links, including the entrance and exit links, are totally jammed, and no vehicles can move around. This can occur when all exit links are blocked. However, more interesting.

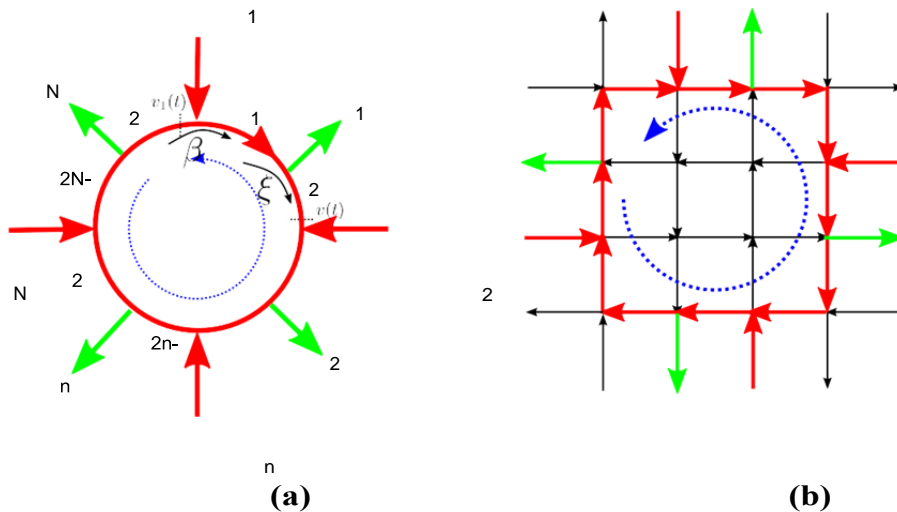


FIGURE 1 Beltways embedded in (a) a freeway network and (b) an arterial grid network.

gridlock states can arise when all entrance links are jammed but all exit links are completely empty. The latter type of nontrivial gridlock states are concerned in this study

In Daganzo (1) it was shown that a nontrivial gridlock state can appear in a beltway network, and the sufficient condition for the occurrence of gridlock in such networks was identified with macroscopic merging and diverging rules in the cell transmission model (2). In Daganzo (3) the occurrence of gridlock was also demonstrated in an urban beltway network with continuous entrance and exit links. Further, in Daganzo (4) the dynamics related to the occurrence and mitigation of gridlock were studied in double-ring networks, and the impacts of drivers' adaptation to traffic conditions were analyzed. In Jin et al. (5) it was shown that a double-ring network can reach a gridlock state at multiple density levels in the kinematic wave model. In Jin (6) it was shown that the occurrence of a gridlock state is associated with circular information propagation on a ring road, and that a gridlock state is a fixed point of a Poincaré map. In the literature, however, there lacks a systematic treatment on existence, stability, and mitigation of gridlock in a beltway network.

This study attempts to fill the gap. First the network kinematic wave model, in which the Lighthill-Whitham-Richards (LWR) model is used to describe traffic dynamics on each link, is applied, and macroscopic merging and diverging rules are used to prescribe traffic dynamics at merging and diverging junctions. Then we mathematically prove that, in a traffic statics problem with constant demands at the entrance links and constant supplies at the exit links, there exist stationary states in the network, and gridlock is always a stationary solution. Furthermore, with the Poincaré map developed in Jin (6), this paper shows that the gridlock state may be stable or unstable, depending on relationship between the merging priority and the diverging ratio. Finally, a feedback control algorithm to mitigate gridlock is proposed, and its effect on the existence and stability of gridlock is discussed.

The rest of the paper is organized as follows. In Section 2 the network kinematic wave model for traffic dynamics in the beltway network is presented. In Section 3 the traffic statics problem in the beltway network is formulated, and the existence of gridlock stationary states is proved. In Section 4 the Poincaré map is applied to study the stability of the gridlock state. In

Section 5, the impacts of an evacuation diverging model on the gridlock state are studied. In Section 6, the paper concludes with discussions and possible future studies.

NETWORK KINEMATIC WAVE MODEL

For a rotationally symmetric beltway network, shown in Figure 1a, N entrances (or on-ramps) and N exits (or off-ramps) ($N \geq 2$) are labeled from 1 to N , respectively. It is assumed that all mainline links between an entrance and an exit have the same length, L , and they are labeled from 1 to $2N$ in a clockwise fashion. In addition, the merging junctions are labeled by $2n - 1$, and the diverging junctions are labeled by $2n$ ($n = 1, \dots, N$).

On mainline link a ($a = 1, \dots, 2N$), at a point (a, x_a) and time t , traffic density, speed, and flow-rate are denoted by $k_a(x_a, t)$, $v_a(x_a, t)$, and $q_a(x_a, t)$, respectively. It is assumed that all links share the same fundamental diagram (7): $q = Q(k)$, which is unimodal and attains its capacity $C = Q(K_c)$ at the critical density of K_c . An example is the following triangular fundamental diagram:

$$Q(k_a) = \min\{V k_a, W(K - k_a)\} \quad (1)$$

where V is the free-flow speed, K is the jam density, and $-W$ is the shockwave speed in congested traffic. For a flow-rate q_a , $K_1(q_a)$ and $K_2(q_a)$ are the respective undercritical (UC) and overcritical (OC) densities for q_a ; in other words, $Q(K_1(q_a)) = Q(K_2(q_a)) = q_a$, and $K_1(q_a) \leq K_c \leq K_2(q_a)$. The triangular fundamental diagram and $K_1(\cdot)$ and $K_2(\cdot)$ are shown in Figure 2.

Traffic dynamics on link a , either a mainline, entrance, or exit link, are described by the LWR model (8, 9):

$$\frac{\partial k_a}{\partial t} + \frac{\partial Q(k_a)}{\partial x_a} = 0 \tag{2}$$

in which traffic queues propagate and dissipate with shockwaves. In the celebrated cell transmission model (2, 10) the LWR model is extended for traffic dynamics in a network in the following two steps. First, the following traffic demand and supply functions are defined:

$$d = D(k_a) \equiv Q(\min\{K_c, k_a\}), \text{ and} \tag{3a}$$

$$s = S(k_a) \equiv Q(\max\{K_c, k_a\}). \tag{3b}$$

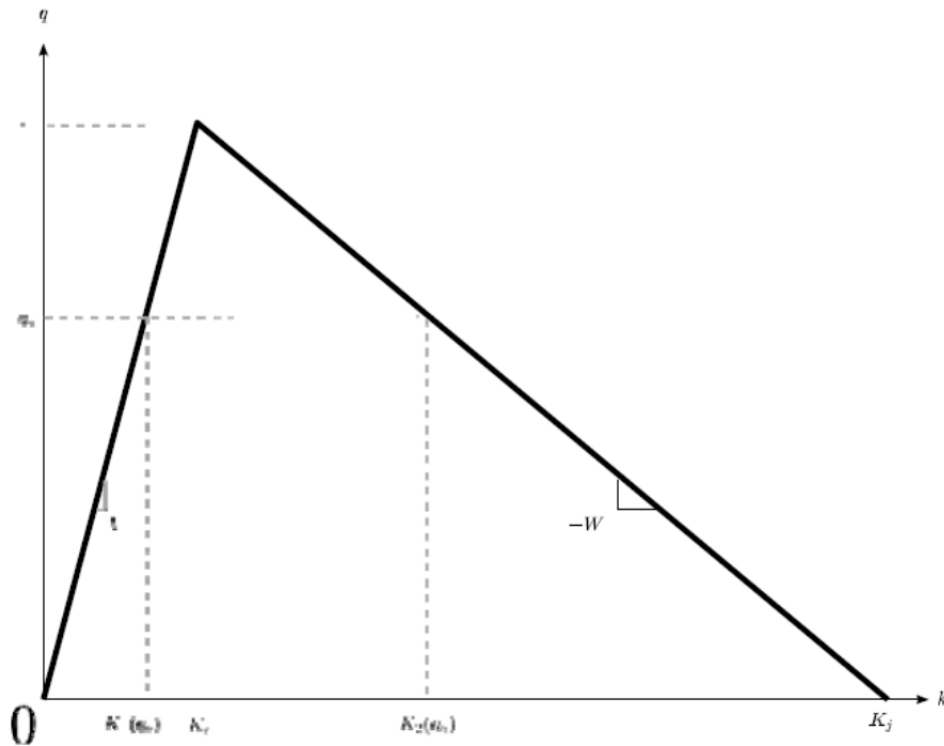


FIGURE 2 The triangular fundamental diagram and flow–density relation.

Second, traffic dynamics at the merging and diverging junctions can be described by the following macroscopic models:

(a) At merging junction $2n - 1$, the boundary fluxes are given by

$$q_{2n-1}(0, t) = \min\{d_{2n-2}(L^-, t) + \delta_n(t), s_{2n-1}(0^+, t)\}, \quad (4a)$$

$$q_{2n-2}(L, t) = \min\{d_{2n-2}(L^-, t), \max\{s_{2n-1}(0, t) - \delta_n(t), \beta s_{2n-1}(0, t)\}\} \quad (4b)$$

where $2n - 2 = 2N$ when $n = 1$, $q_{2n-1}(0, t)$ is the influx of link $2n - 1$, $q_{2n-2}(L, t)$ the out-flux of link $2n - 2$, $d_{2n-2}(L^-, t)$ the downstream demand of link $2n - 2$, $s_{2n-1}(0, t)$ the upstream supply of link $2n - 1$, and $\delta_n(t)$ the demand of entrance n . Here $1 - \beta$ is the merging ratio of the entrance, and β that of the mainline link. This is the priority-based merging model proposed in Daganzo (2), which was shown to be invariant in Jin (11) and thus can be used to analyze traffic dynamics in the network without worrying about the existence of so-called interior states.

(b) At diverging junction $2n$, the boundary fluxes are given by

$$q_{2n-1}(L, t) = \min\left\{d_{2n-1}(L^-, t), \frac{s_{2n}(0^+, t)}{\xi}, \frac{\sigma_n(t)}{1-\xi}\right\} \quad (5a)$$

$$q_{2n}(0, t) = \xi q_{2n-1}(L, t) \quad (5b)$$

where $q_{2n-1}(L, t)$ is the out-flux of link $2n - 1$, $q_{2n}(0, t)$ is the influx of link $2n$, $d_{2n-1}(L^-, t)$ is the downstream demand of link $2n - 1$, $s_{2n}(0^+, t)$ is the upstream supply of link $2n$, and $\sigma_n(t)$ is the supply of exit n . Here $1 - \xi$ is the diverging proportion to the exit, and ξ that to the mainline link. This is the first-in–first-out diverging model proposed in Daganzo (2) and was shown to be invariant in Jin (12). Note that here all vehicles are assumed to have predefined destinations in the diverge model (5).

Thus, Equations 2, 4, and 5 form the network kinematic wave model, which is the continuous version of the cell transmission model, for traffic dynamics in the beltway network. It can be considered a control system, where the traffic densities, $k_a(x_a, t)$, are state variables; the entrance demands, $\delta_n(t)$, and exit supplies, $\sigma_n(t)$, are inputs; and the merging priority, β , and the diverging ratio, ξ , are control variables. Note that $k_a(x_a, t)$ is of an infinite dimension, and the control system is infinite-dimensional and cannot be easily analyzed.

TRAFFIC STATICS PROBLEM AND EXISTENCE OF GRIDLOCK

Following Jin (13), the traffic statics problem for the beltway network is defined as finding stationary solutions to the kinematic wave model, Equations 2, 4, and 5, when the origin demands and destination supplies are all constant; in other words, $\delta_n(t) = \delta$ and $\sigma_n(t) = \sigma$.

Stationary Traffic Conditions on Links and at Junctions

As shown in Jin (5, 13), when link a becomes stationary, its density can be written as ($x_a \in [0, L]$)

$$k_a(x_a, t) = H(u_a L - x_a)K_1(q_a) + (1 - H(u_a L - x_a))K_2(q_a) \quad (6)$$

where $u_a \in [0, 1]$ is the uncongested fraction of the road, $H(\cdot)$ is the Heaviside function.

The upstream supply and the downstream demand are both constant: $s_a(0^+, t) = s_a$, and $d_a(L^-, t) = d_a$. Thus there are four types of stationary states:

$$\text{Strictly undercritical: } u_a = 1, q_a < C, d_a = q_a, \text{ and } s_a = C \quad (7a)$$

$$\text{Strictly overcritical: } u_a = 0, q_a < C, d_a = C, \text{ and } s_a = q_a \quad (7b)$$

$$\text{Critical (C): } u_a \in [0, 1], q_a = C, d_a = C, \text{ and } s_a = C \quad (7c)$$

$$\text{Zero-speed shockwave: } u_a \in (0, 1), q_a < C, d_a = C, \text{ and } s_a = C \quad (7d)$$

A stationary state on link a is illustrated in Figure 3. Note that all types of stationary states can be considered as special cases of a zero-speed shockwave. In stationary states around merging and diverging junctions, Equations 4 and 5 can be simplified as follows:

1. At merging junction $2n - 1$,

$$q_{2n-1} = \min\{d_{2n-2} + \delta, s_{2n-1}\} \quad (8a)$$

$$q_{2n-2} = \min\{d_{2n-2}, \max\{s_{2n-1} - \delta, \beta s_{2n-1}\}\} \quad (8b)$$

2. At diverging junction $2n$,

$$q_{2n-1} = \min\left\{d_{2n-1}, \frac{s_{2n}}{\xi}, \frac{\sigma}{1-\xi}\right\} \quad (9a)$$

$$q_{2n} = \xi q_{2n-1} \quad (9b)$$

Therefore, to solve the traffic statics problem, it is necessary to find q_a , u_a , d_a , and s_a for $a = 1, \dots, 2N$ from the system of algebraic equations in Equations 7, 8, and 9. In general, since each link can have four possible types of stationary states, there can be 4^{2N} possible combinations, and a brute force method as in Jin (13) cannot apply in this case.

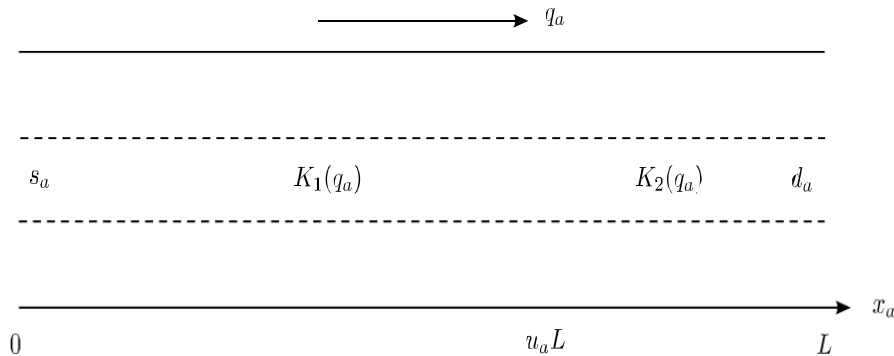


FIGURE 3 Stationary states on a link.

Existence of Gridlock and Other Stationary Solutions

In the gridlock state in the beltway network, there are $q_a = s_a = 0$, $u_a = 1$, and $d_a = C$ for $a = 1, \dots, 2N$.

Lemma 3.1. If $\delta > 0$ and $\sigma > 0$, then the entrance links are jammed, and the exit links are empty.

Proof. For entrance link n at merging junction $2n - 1$, since the flow-rates on links $2n - 1$ and $2n - 2$ are both 0, the flow-rate on the entrance link is also 0. Thus its upstream supply, s , has to be 0, since, otherwise, its upstream flow-rate equals $\min\{\delta, s\} > 0$. From Equation 7 it is evident that a link whose upstream supply is 0 has to be totally jammed.

For exit link n at diverging junction $2n$, its flow rate is also 0. Thus its downstream supply, d , has to be 0, since, otherwise, its downstream flow rate equals $\min\{d, \sigma\} > 0$. From Equation 7 it appears that a link whose upstream supply is 0 has to be empty.

Theorem 3.2. In a beltway network, the gridlock state is always a stationary solution.

Proof. First, clearly the gridlock state is a stationary state on a link as defined in Equation 7. Second, at merging junction $2n - 1$, since $d_{2n-2} = C$ and $s_{2n-1} = 0$, we have

$$q_{2n-1} = 0 = \min\{C + \delta, 0\}$$

$$q_{2n-2} = 0 = \min\{C, \max\{-\delta, 0\}\}$$

Thus, the gridlock state satisfies the merge model. Third, at diverging junction $2n$,

since $d_{2n-1} = 0$ and $s_{2n} = 0$, we have

$$q_{2n-1} = 0 = \min\left\{C, 0, \frac{\sigma}{1 - \xi}\right\}$$

$$q_{2n} = 0 = \xi q_{2n-1}$$

Thus, the gridlock state satisfies the diverge model. Therefore the gridlock state is a stationary solution of the network kinematic wave model with constant entrance demand and exit supply.

Theorem 3.2 is a very important observation, even though the mathematical proof is quite straightforward. From this theorem one can conclude that, if a beltway network is initially jammed, it will always be jammed. Given Lemma 3.1, this is a bit surprising, since there are sufficient spaces at the exit links and sufficient vehicles at the entrance links. The occurrence of such a gridlock state is highly related to the diverging model, in which all vehicles have predefined routes, even though there are better options in other routes. This suggests that route guidance strategies can be helpful for preventing the occurrence of gridlock in a beltway network.

STABILITY OF GRIDLOCK

In the gridlock state, all mainline links are congested, and traffic waves propagate upstream. Thus, there is a counterclockwise circular information propagation path along the ring road. As shown in Figure 1, two Poincaré sections are defined at the upstream points of two consecutive merging junctions 3 and 1, and denote the two out-fluxes of the mainline road by $v(t) = q_2(L, t)$ and $v_1(t) = q_{2N}(L, t)$, respectively.

After T_1 , $v(t)$ propagates to the diverging junction 2, and $s_2(0^+, t + T_1) = v(t)$, since link 2 is congested. From Equation 5 there is

$$q_1(L, t + T_1) = \frac{v(t)}{\xi}$$

After T_2 , $q_1(L, t + T_1)$ propagates to the merging junction 1, and $s_1(0^+, t + T_1 + T_2) = q_1(L, t + T_1)$ since link 1 is congested. In Equation 4

$$q_{2N}(L, t + T) = \beta q_1(L, t + T_1) = \frac{\beta}{\xi} v(t)$$

where $T = T_1 + T_2$. Then, after N pairs of entrance and exit links, the following Poincaré map is obtained:

$$v(t + NT) = \left(\frac{\beta}{\xi}\right)^N v(t) \quad (10)$$

Note that the Poincaré map is valid only when all links are congested and cannot capture other types of stationary states.

For the Poincaré map, (Equation 10), the gridlock state $v(t) = 0$ is always an equilibrium point. Furthermore, from the Poincaré map, the stability of the gridlock state can be determined:

1. When $\frac{\beta}{\xi} < 1$, the gridlock state is asymptotically stable, and the half-life equals $\frac{\ln 2}{-N \ln(\frac{\beta}{\xi})}$.
2. When $\frac{\beta}{\xi} > 1$, the gridlock state is unstable.
3. When $\frac{\beta}{\xi} = 1$, there can exist multiple Lyapunov stable stationary states. It can be verified that these results are consistent with those in Daganzo (1).

MITIGATION OF GRIDLOCK WITH ADAPTIVE DRIVING

Existing Mitigation Strategies

For a beltway network described by Equation 2 with Equations 4 and 5, when the exit links are uncongested, three types of control strategies have been discussed in Daganzo (1, 3, 4) to mitigate the impacts of gridlock:

1. The demand $\delta_n(t)$ can be regulated by ramp metering. This may delay the occurrence of gridlock. But when all links are congested, if $\frac{\beta}{\xi} < 1$, the system will converge to the gridlock state irreversibly.
2. The merging ratio for the entrances, $1 - \beta$, can be decreased by signal settings at the intersections. This can destabilize the gridlock state when $\frac{\beta}{\xi} > 1$.
3. The diverging proportion to the exits, $1 - \zeta$, can be increased by variable message signs and adaptive driving behaviors. This can also destabilize the gridlock state when $\frac{\beta}{\xi} > 1$.

These results can be explained by Equation 10. However, as shown in Section 3, these strategies may not avoid the occurrence of gridlock, since the gridlock state is always a stationary solution to the system.

A New Adaptive Driving Strategy

This subsection addresses another type of adaptive driving strategy, derived from the evacuation model in Jin (5). At a diverge $2n$, the following diverging model is applied:

$$q_{2n-1}(L, t) = \min\{d_{2n-1}(L^-, t), s_{2n}(0^+, t) + \sigma_n(t)\} \quad (11a)$$

$$q_{2n}(0, t) = \min\{s_{2n}(0^+, t), \max\{d_{2n-1}(L^-, t) - \sigma_n(t), \xi d_{2n-1}(L^-, t)\}\} \quad (11b)$$

where $\xi \in [0, 1]$ is the priority for adaptive drivers to choose the ring road. In a sense, this is a feedback control strategy, as the diverging ratio is determined by the dynamic traffic conditions.

Consider symmetric stationary solutions on the ring road, in which all odd-numbered links share the same traffic conditions, denoted by q_1 , u_1 , d_1 , and s_1 , and all even-numbered mainline links share the same traffic conditions, denoted by q_2 , u_2 , d_2 , and s_2 . Thus, from Equations 8 and 11

$$q_1 = \min\{d_2 + \delta, s_1\} \quad (12a)$$

$$q_2 = \min\{d_2, \max\{\sigma_1 - \delta, \beta s_1\}\} \quad (12b)$$

$$q_1 = \min\{d_1, s_2 + \sigma\} \quad (12c)$$

$$q_2 = \min\{s_2, \max\{d_1 - \sigma, \xi d_1\}\} \quad (12d)$$

Lemma 5.1. When $\delta > 0$ and $\sigma > 0$, the gridlock state is no longer a solution of Equation 12.

Proof. It is assumed that the gridlock state is still a stationary solution of Equation 12; in other words, $d_2 = d_1 = C$, and $s_1 = s_2 = 0$. From the first two equations, it is clear that $q_1 = q_2 = 0$. However, from the third equation, $0 = \min\{C, \sigma\} > 0$, which is impossible. Thus, the gridlock state is no longer a solution.

Note that the adaptive driving strategy yields the same results as without using it under uncongested conditions. In this case, $q_1 = d_1 \leq s_1 = C$, and from (12c) and (12d) comes $q_1 = d_1$ and $q_2 = \xi d_1$. That is, vehicles can still use their original routes under uncongested conditions. However, when traffic is congested, some vehicles will have to be rerouted to avoid the development of gridlock.

CONCLUSION

For this study researchers formulated the traffic dynamics in a beltway network with the network kinematic wave model incorporating the LWR model as well as invariant merging and diverging models. Then, stationary states, including the gridlock state, and their stability, were studied. Finally, existing mitigation strategies were discussed, and a new evacuation-based strategy that eliminates the gridlock state in stationary states was proposed.

The closed-form network kinematic wave model of the traffic dynamics in the beltway network has enabled the analyses of stationary states and their stability properties. In this study, however, the discussions are at the macroscopic level. In the future researchers will be interested in finding the optimal strategy that prevents gridlock but involves the minimum number of reroutings at the microscopic level. In addition, researchers will be interested in designing the best mitigation strategies that can lead to the minimum delays and in designing the adaptive control strategies for mitigating traffic congestion and avoiding gridlock in more general networks.

REFERENCES

1. Daganzo, C. F. The Nature of Freeway Gridlock and How to Prevent It. *Proc., 13th International Symposium on Transportation and Traffic Theory*, 1996, pp. 629–646.
2. Daganzo, C. F. The Cell Transmission Model II: Network Traffic. *Transportation Research Part B*, Vol. 29, No. 2, 1995, pp. 79–93.
3. Daganzo, C. F. Urban Gridlock: Macroscopic Modeling and Mitigation Approaches. *Transportation Research Part B*, Vol. 41, No. 1, 2007, pp. 49–62.
4. Daganzo, C. F. On the Macroscopic Stability of Freeway Traffic. *Transportation Research Part B*, Vol. 45, No. 5, 2011, pp. 782–788.
5. Jin, W.-L., Q.-J. Gan, and V. V. Gayah. A Kinematic Wave Approach to Traffic Statics and Dynamics in a Double-Ring Network. *Transportation Research Part B*, Vol. 57, 2013, pp. 114–131.
6. Jin, W.-L. Stability and Bifurcation in Network Traffic Flow: A Poincaré Map Approach. *Transportation Research Part B*, Vol. 57, 2013, pp. 191–208.
7. Greenshields, B. D., J. R. Bibbins, W. S. Channing, and H. H. Miller. A Study of Traffic Capacity. Presented at 14th Annual Meeting of the Highway Research Board, Washington, D.C., 1935.
8. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves: II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society of London A*, Vol. 229, No. 1178, 1955, pp. 317–345.
9. Richards, P. I. Shock Waves on the Highway. *Operations Research*, Vol. 4, No. 1, 1956, pp. 42–51.

10. Lebacque, J. P. The Godunov Scheme and What It Means for First Order Traffic Flow Models. *Proc., 13th International Symposium on Transportation and Traffic Theory*, 1996, pp. 647–678.
11. Jin, W.-L. Continuous Kinematic Wave Models of Merging Traffic Flow. *Transportation Research Part B*, Vol. 44, No. 8–9, 2010, pp. 1084–1103.
12. Jin, W.-L. Analysis of Kinematic Waves Arising in Diverging Traffic Flow Models. *Transportation Science*, 2014, <http://arxiv.org/abs/1009.4950>
13. Jin, W.-L. The Traffic Statics Problem in a Road Network. *Transportation Research Part B*, Vol. 46, No. 10, 2012, pp. 1360–1373.

NETWORKWIDE MODELING AND CONTROL

Effects of Segregating Buses and Cars in a Congested, Non–Steady-State Street Network

NATHALIE SAADE

MICHAEL J. CASSIDY

University of California, Berkeley

WEIHUA GU

Hong Kong Polytechnic University

Much of the literature on exclusive bus lanes pertains to how these lanes might induce shifts in mode choice by prioritizing bus travel, sometimes at the expense of degrading travel by car; for example, see Basso et al. (1). A separate line of research theorized that the conversion of regular-use lanes to bus-only lanes can, in certain circumstances, improve travel for cars as well as buses, even in the absence of modal shifts (2). By removing buses from queues and putting them in their own, faster-moving lanes, target service frequencies can be maintained with fewer buses; thus, fewer bus lanes are needed and more lanes can be left for the exclusive use of cars.

However, these Pareto improvements were predicted in Cassidy et al. (2) only when converted lanes enjoyed rather high bus flows and only by relying on a so-called “smoothing effect,” which describes the network capacity gained when distinct travel modes are segregated into their own lanes. With this segregation, disruptive vehicular interactions are diminished, which can generate higher bottleneck capacities (3, 4). In efforts to garner the kinds of high-level insights that can guide large-scale planning decisions, street networks were modeled in Cassidy et al. (2) as rotationally-symmetric, closed-loop beltways operating in the steady-state.¹

The present work follows lines of thought that are similar to those in Cassidy et al. (2), in that it, too, models the impacts of converting regular lanes to bus lanes on rotationally-symmetric beltways and assumes that travel demands for buses and cars are not subject to shifts from one mode to the other. However, the present work recognizes that rush periods are invariably characterized by non–steady-state conditions. In other words, the early part of a rush period is typically characterized by travel demand that exceeds beltway capacity, such that beltway queues expand, while the later part is characterized by demand that falls below capacity, such that queues gradually disappear. It turns out that these non-steady-state realities can be favorable to bus lanes, meaning that lane conversions can sometimes improve travel for buses and cars, even if one ignores the smoothing effect.

ASSUMPTIONS AND METHODS

This paper assumes a set of conditions that might describe the operation of jitney buses in a city within the developing world, albeit in idealized fashion. Contrary to Cassidy et al. (2), it is assumed that the number of bus trips made during the rush is fixed, irrespective of whether some lanes are given to buses. Each bus enters a beltway network during the rush, serves assigned portions of the network, and then exits. The points from which buses and cars enter the beltway

are uniformly distributed over its boundary. When buses and cars share the same lanes, both modes compete for available road capacity to enter and circulate through the beltway. When some lanes are instead set aside for buses, those buses enter and circulate without delay, and cars are left with fewer lanes on the beltway for their use. The number of bus lanes to be used on the beltway, N , must be integer valued such that $1 \leq N \leq NT - 1$, where NT is the beltway's total number of lanes available to serve traffic in each travel direction. The upper bound guarantees that at least one lane is available in each direction to serve cars. Very importantly, once all rush-period bus tours are completed, all NT lanes are thereafter made available to cars.

It is further assumed that prior to the start of a rush at time $t = 0$, there is a fixed demand in passenger car equivalents (pce) that is less than beltway capacity; from $t = 0$ to $t = tr$, there is a fixed demand that exceeds the previous rate by the amount λ (in pce/h), and this high demand exceeds beltway capacity, and at $t > tr$, demand returns to the initial low rate that occurred prior to $t = 0$.

For simplicity, the average speed and the average distance traveled on the beltway are all measured by pce. The implicit assumption is that the variances in speed and in distance traveled are not large, neither when measured across vehicles of the same class nor across the two distinct classes.

Beltway traffic is described using a triangular-shaped macroscopic fundamental diagram (MFD) that relates total flow on the network (in pce/h) to network-wide density (pce/km), much as in Daganzo (5). Knowing the beltway's physical size and the vehicle trip length on it, the MFD is rescaled to a network exit function (NEF) that relates the total vehicle accumulation on the beltway, n (pce), to the rate that vehicles complete their trips on it, f (pce/h). The maximum trip-completion rate is f_m and the corresponding accumulation is n_m .

For this paper the beltway network is viewed as being composed of a series of identical building blocks, like the building block shown for a single travel direction in Figure 1. When the beltway is congested $a_j > \beta_j$, where a_j is the proportion of pce that enters the block j from outside the beltway, and β_j is the proportion of the block's pce that exits the beltway. Hence for every pce that enters a block via the beltway, there are $(1 + a_j)$ that traverse the block and remain on the beltway. Given that the blocks are identical and that the beltway is rotationally symmetric, a_j and β_j are aggregated across all the beltway blocks. Hence, $a = \sum_j a_j$ and $\beta = \sum_j \beta_j$ represent the proportion of pce that enters and exits the beltway in total. The merge model of Daganzo (5) is used to estimate the block entry flows at short time steps.

With entry and exit flows thusly obtained for a point in time, t , the NEF is used to determine the accumulation in the beltway, $n(t)$. Then, a queueing diagram like the one in Figure 2 can be incrementally constructed to estimate delays.

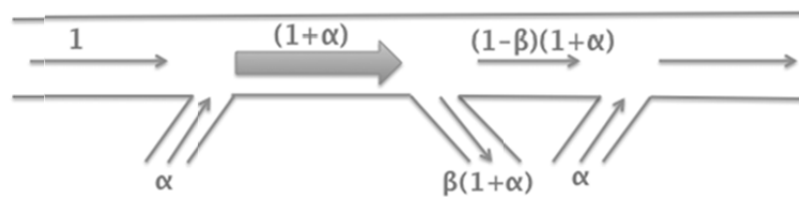


FIGURE 1 Building block of the beltway.

The diagram in Figure 2 was constructed for the case in which buses and cars share the same lanes. (The fixed flow at $t < 0$ and $t > t_r$ is treated as a background rate and was subtracted from the cumulative count curves.) When bus lanes are used to segregate buses and cars, the analysis is performed using the demand for car travel (only), since buses are no longer subject to delays. The merge model and the NEF are rescaled to account for the diminished space that remains available to car traffic.

Parametric analysis featuring comparisons of segregated and nonsegregated cases unveils certain insights. Examples follow.

EXAMPLE FINDINGS

Figure 3 presents analysis outcomes for ranges of t_r (normalized by average vehicle free-flow trip time on the network) and λ (normalized by f_m). Other key inputs to the analysis are noted in the figure caption. Most of these were borrowed from Geroliminis and Daganzo (6).

The boundary lines near the graph's upper-right corner delineate where combinations of t_r and λ create gridlock on the beltway network. Tellingly, the gridlock region is smaller for the segregated case, where one or more beltway lanes are given over to buses.

The darker-shaded area at the bottom of the graph highlights the combinations of t_r and λ , for which bus lanes increase the person-hours traveled by car on the network, and even the total person-hours traveled by both modes combined, as compared against not setting aside any lanes for bus-use only. The lighter-shaded region directly atop denotes combinations for which the bus lanes still increase the person-hours traveled by car but diminish the total hours traveled via both modes combined. Very importantly, the large unshaded portion of the graph shows where bus lanes diminish not only the combined total hours traveled but also the hours traveled by cars.

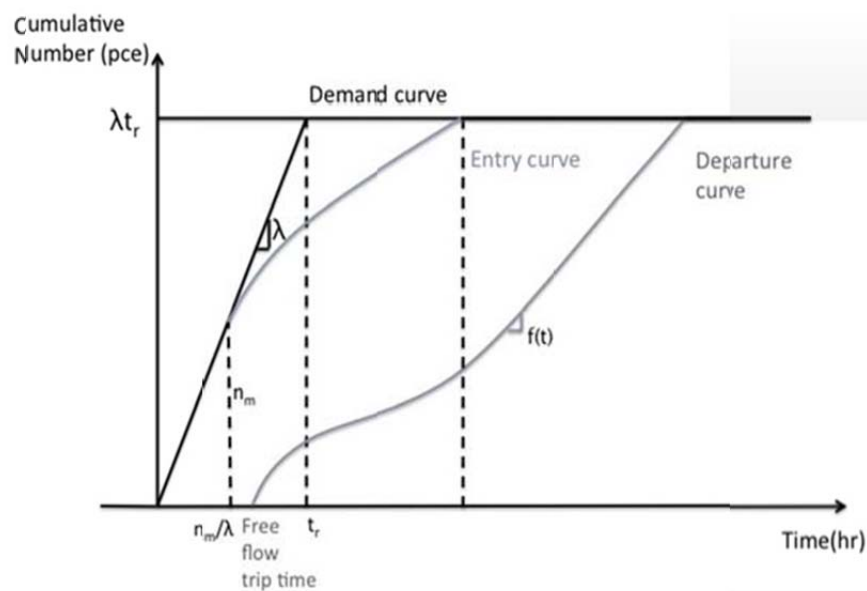


FIGURE 2 Cumulative total demand, entry and departure curves.

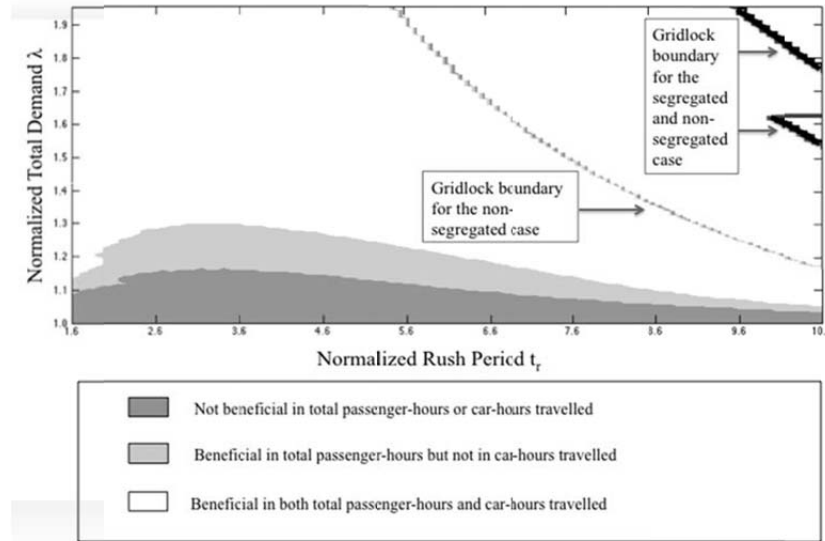


FIGURE 3 Delay reduction map for a bus demand proportion of 20%, a number of lanes of 3, $f_m = 15,660$ pce/h, $n_m = 3,000$ pce, $a = 0.2$, $p = 0.14$, a maximum on-ramp capacity of $2f_m$, and a maximum total accumulation of 6,500 pce.

Car-hours diminish in these latter cases because by enabling buses to travel faster, the bus lanes mean that the total collection of bus tours will end earlier in the rush. Thereafter, all lanes are given over for car travel, and cars no longer compete with rush-period buses for beltway capacity.²

The findings are even more favorable to bus lanes when one considers the smoothing effect.

Other illustrative analyses and other insights are offered in the full paper.

NOTES

1. The beltway is a good representation of a congested downtown street network. For example, the beltway's on- and off-ramps can represent the downtown's access and egress points; the shorter the downtown's block lengths, the greater the number of beltway ramps. And the beltway's circumference can be selected so as to account for the downtown's total available road space.

2. The bus lanes are not Pareto improving in these cases, because car travelers during the early part of the rush incur greater delays owing to the fewer number of beltway lanes available to them. Happily, these added delays are more than compensated for by the delay savings that cars enjoy later in the rush.

REFERENCES

1. Basso, L. J., C. A. Guevara, A. Gschwender, and M. Fuster. Congestion Pricing, Transit Subsidies and Dedicated Bus Lanes: Efficient and Practical Solutions to Congestion. *Transport Policy*, Vol. 18, No. 5, 2011, pp. 676–684.
2. Cassidy, M. J., C. F. Daganzo, K. Jang, and K. Chung. Spatiotemporal Effects of Segregating

- Different Vehicle Classes on Separate Lanes. In *Transportation and Traffic Flow Theory 2009: Golden Jubilee* (W. H. Lam, S. C. Wong, and H. K. Lo, eds.), Springer, 2009, pp. 57–74.
3. Menendez, M., C. F. Daganzo. Effects of HOV Lanes on Freeway Bottleneck. *Transportation Research Part B*, Vol. 41, 2007, pp. 809–822.
 4. Cassidy, M. J., K. Jang, and C. F. Daganzo. The Smoothing Effect of Carpool Lanes on Freeway Bottlenecks. *Transportation Research Part A: Policy and Practice*, Vol. 44, No. 2, 2010, pp. 65–75.
 5. Daganzo, C. F. The Nature of Freeway Gridlock and How to Prevent It. *Proc., 13th International Symposium on Transportation and Traffic Theory*, 1996, pp. 629–646.
 6. Geroliminis, N., and C. F. Daganzo. Macroscopic Modeling of Traffic in Cities. Presented at 86th Annual Meeting of the Transportation Research Board, Washington D.C., 2007.

NETWORKWIDE MODELING AND CONTROL

Stochastic Approximations for the Macroscopic Fundamental Diagram

JORGE A. LAVA

FELIPE CASTRILLÓN

YI ZHOU

Georgia Institute of Technology

This paper proposes analytical estimates of a network macroscopic fundamental diagram (MFD) using probabilistic methods. It produces a distribution for the MFD, found to depend mainly on two parameters. This abstract focuses on [Figure 1](#), which shows the comparison between the MFDs from an exact traffic simulation and the 90% probability interval from the authors' analytical method.

Existing analysis methods for the MFD fall into three types: 1. empirical (1–12), 2. analytical (13–15), and 3. simulation (16–22). Analytical results are based on the method of cuts for homogeneous corridors where all intersections see the same set of cuts, so the cuts from a single intersection suffice to compute the MFD for the whole corridor. Despite this apparent simplicity, this approach quickly becomes intractable, and one needs to resort to simple but numerical algorithms to identify all possible cuts. Additionally, the homogeneous corridor method cannot be scaled up without complications to estimate the network MFD, mainly because a network path cannot be guaranteed to have constant offset all along, even for homogeneous networks.

This paper introduces the concept of probabilistic homogeneous corridors, where all intersections see the same distribution of cuts. This allows tackling deterministic inhomogeneous corridors with different block length, signal timing, and others.

The example presented in [Figure 1](#) is of networks with random offsets; the general method will be presented in the final paper. The general method can be used to randomize other variables of the problem such as block length and signal timing; this is important because it allows the analytical estimation of MFDs on inhomogeneous networks, for which analytical methods are currently unavailable.

Random offset networks could be a good approximation for networks with traffic actuated signals or without coordination, and [Figure 1](#) shows that it provides tight bounds for the deterministic problem. The dimensionless formulation of this problem provides new insights, including the following examples: 1. Only two parameters determine the MFD distribution, 2. The distribution of network capacity depends on a single parameter, 3. The existence of a symmetry line in the MFD, which is crucial for mathematical tractability, and 4. The block length does not affect the MFD when it is larger than twice the critical block length.

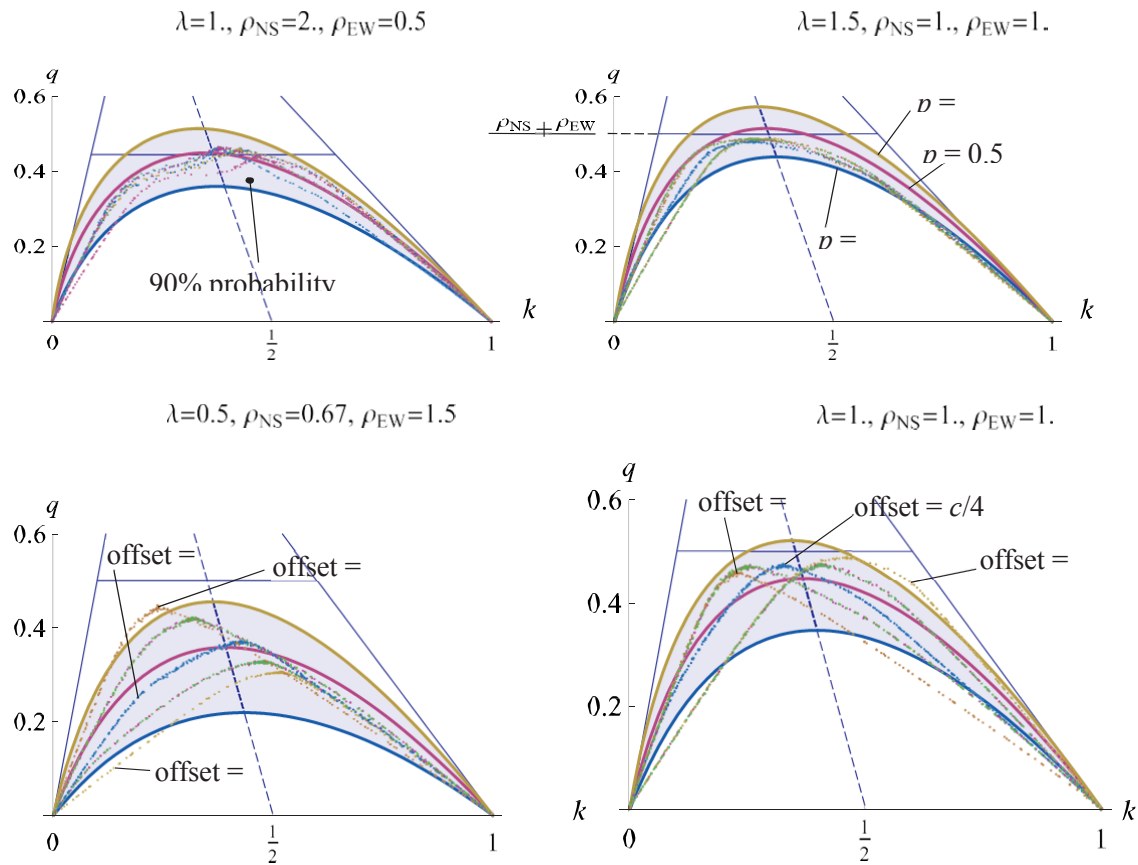


FIGURE 1 Comparison between the MFDs produced by an exact traffic simulation and the 90% probability interval produced by the analytical method.

REFERENCES

1. Geroliminis, N., and C. Daganzo. Macroscopic Modeling of Traffic in Cities. Presented at 86th Annual Meeting of the Transportation Research Board, Washington, D.C., 2007.
2. Geroliminis, N., and C. Daganzo. Existence of Urban-Scale Macroscopic Fundamental Diagrams: Some Experimental Findings. *Transportation Research Part B: Methodological*, Vol. 42, 2008, pp. 759–770.
3. Wu, X., H. Liu, and N. Geroliminis. An Empirical Analysis on the Arterial Fundamental Diagram. *Transportation Research Part B: Methodological*, Vol. 45, No. 1, 2011, pp. 255–266.
4. Saberi, M., and H. S. Mahmassani. Exploring the Properties of Networkwide Flow–Density Relations in a Freeway Network. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2315, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 153–163.
5. Helbing, B. D. Derivation of a Fundamental Diagram for Urban Traffic Flow. *The European Physical Journal B*, Vol. 70, 2009, pp. 229–241.
6. Geroliminis, N., and J. Sun. Properties of a Well-Defined Macroscopic Fundamental Diagram for Urban Traffic. *Transportation Research Part B*, Vol. 45, 2011, pp. 605–617.
7. Geroliminis, N., and Y. Ji. Spatial and Temporal Analysis of Congestion in Urban Transportation Networks. Presented at 90th Annual Meeting of the Transportation Research Board, Washington D.C., 2011.

8. Cassidy, M., K. Jang, and C. Daganzo. Macroscopic Fundamental Diagram for Freeway Networks: Theory and Observation. Presented at 90th Annual Meeting of the Transportation Research Board, Washington D.C., 2011.
9. Knoop, V. L. The Impact of Traffic Dynamics on the Macroscopic Fundamental Diagram. Presented at 92th Annual Meeting of the Transportation Research Board, Washington D.C., 2012.
10. Gayah, V., and C. Daganzo. Clockwise Hysteresis Loops in the Macroscopic Fundamental Diagram: An Effect of Network Instability, *Transportation Research Part B*, 2011.
11. Buisson, C., and C. Ladier. Exploring the Impact of Homogeneity of Traffic Measurements on the Existence of Macroscopic Fundamental Diagrams. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2124*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 127–136.
12. Daganzo, C. F., V. V. Gayah, and E. J. Gonzales. Macroscopic Relations of Urban Traffic Variables: Bifurcations, Multi-Valuedness and Instability. *Transportation Research Part B: Methodological*, Vol. 45, 2011, pp. 278–288.
13. Daganzo, C. F., and N. Geroliminis. An Analytical Approximation for the Macroscopic Fundamental Diagram of Urban Traffic. *Transportation Research Part B: Methodological*, Vol. 42, 2008, pp. 771–781.
14. Daganzo, C. F. A Variational Formulation of Kinematic Waves: Basic Theory and Complex Boundary Conditions. *Transportation Research Part B*, 39(2), 2005, pp. 187–196.
15. Daganzo, C. F. A Variational Formulation of Kinematic Waves: Solution Methods. *Transportation Research Part B*, Vol. 39, No. 10, 2005, pp. 934–950.
16. Ji, Y., W. Daamen, S. Hoogendoorn, S. Hoogendoorn-Lanser, and X. Qian. Investigating the Shape of the Macroscopic Fundamental Diagram Using Simulation Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2161*, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 40–48.
17. Mazlomian, A., N. Geroliminis, and D. Helbing. The Spatial Variability of Vehicle Densities as Determinant of Urban Network Capacity. *Philosophical Transactions of the Royal Society*, Vol. A368, 2010, 4627–4647.
18. Geroliminis, N., and B. Boyacı. The Effect of Variability of Urban Systems Characteristics in the Network Capacity. *Transportation Research Part B*, Vol. 46, No. 10, 2013, pp. 1576–1590.
19. Haddad, J., and N. Geroliminis. On the Stability of Traffic Perimeter Control in Two-Region Urban Cities. *Transportation Research Part B*, Vol. 46, No. 9, 2012, pp. 1159–1176.
20. Haddad, J., M. Ramezani, and N. Geroliminis. Cooperative Traffic Control of a Mixed Network with Two Urban Regions and a Freeway. *Transportation Research Part B*, Vol. 54, 2013, pp. 17–36.
21. Knoop, V., and S. Hoogendoorn. Variable Macroscopic Fundamental Diagrams for Traffic Networks. In *Traffic and Granular Flow '11*, Springer, 2011.
22. Knoop, V., J. V. Lint, and S. Hoogendoorn. Route Advice and Its Effect on the Macroscopic Fundamental Diagram. In *Traffic and Granular Flow '11*, Springer, 2011.
23. Daganzo, C. F. In Traffic Flow, Cellular Automata = Kinematic Waves, *Transportation Research Part B*, Vol. 40, No. 5, 2006, pp. 396–403.
24. Lighthill, M. J., and G. Whitham. On Kinematic Waves: I. Flow Movement in Long Rivers. II. A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society of London*, Vol. 229(A), 1955, pp. 281–345.
25. Richards, P. I. Shockwaves on the Highway, *Operations Research*, Vol. 4, 1956, pp. 42–51.

Poster Presentations

POSTER PRESENTATION

Adapting Car Traffic Models and Concepts to Bicycle Traffic

ABDELAZIZ MANAR

GANG CAO

Ville de Montréal, Canada

Over the past 50 years, a huge number of traffic models have been developed to capture and model the driver behaviour on streets, roads and freeways. Almost, all models were developed for motorized modes. Hence, as bicycle use in cities rapidly increases the need for more refined models arises in order to capture and model cyclist behaviour on bike facilities.

In order to capitalize on the years of traffic modeling development, it is legitimate to consider adapting these models to bicycle traffic and thus benefit from all the research efforts in the field of traffic theory. Since, the objective function for a driver, or a cyclist, is to control his vehicle's speed and direction while maintaining his desired speed and avoiding accidents; it is likely, that despite the difference in vehicle type, the driving logic and driver behaviour will be similar. Research has already begun in this area, but compared to what has been done for motorized modes, we are still in the early stages of traffic science for this emerging mode.

This paper presents three aspects of traffic vehicle science adapted to bicycle traffic. These are car-following models, the fundamental relationships of traffic flow and the Action Point Model. The results obtained using car-following models were compared to empirical data collected with global positioning system (GPS) devices installed on a pair of cyclists in a following situation with no opportunity to overtake. The fundamental relationships are examined using data collected by video at a fixed location on a bike facility. Finally, the paper presents a behavioral comparison between the bike-cyclist and the car-driver systems using the Action Point Model.

The results indicate that car-following models, fundamental relationships and the Action Point Model all have the potential to reproduce real world data for bicycle traffic. Also, the commercial simulation software, in this case, VISSIM, can be used with some calibration effort, to reproduce similar results to the observed data.

INTRODUCTION

New realities have emerged with the growing of cycling as a transportation mode in its own right. Thus, cycling is no longer reserved to leisure and its utility now exceeds its recreational use. In addition, occasionally we begin to see on some bicycle facilities in Montréal, Québec, Canada, a phenomenon long reserved to motorized modes: congestion (i.e., situations where platoons of cyclists are forced to follow in single lane given the density of cyclists and lack of overtaking opportunity). This situation can be well modeled by bike-following model.

The need to develop methods and tools to assist engineers and planners to analyze the performance of this mode is increasingly justified (1). Researchers (2, 3) have already begun but compared to what has been done for motorized mode, we are still in the early stages of the traffic science for this emerging mode.

DATA COLLECTION

Two types of data were collected as part of this study in two locations in Montréal. Both locations were off-street, two-way exclusive bicycle facilities.

GPS Data

The first set of data was collected in a controlled experiment. To examine and model the behavior of a pair of cyclists in a strictly following situation, the time–space trajectories of a pair of cyclists on bicycle facility, was recorded by two GPS devices (Garmin and Forerunner 305). The units would provide speed and position for a pair of cyclists every second.

Both devices were synchronized at starting point and secured to wrists of each cyclist, allowing the recording of data on a stretch of 1.7 km on bicycle facility of 1.50 m wide in Montréal. The raw data was then compiled and processed to build the time–space diagram. The time–space diagram is very rich in information, enabling one to calculate an important set of indicators that explain the behavior of the cyclists.

Despite the precision of this GPS device (velocity accuracy <0.05 m/s), some data were cleaned from random errors, attributed essentially to satellite signal blockages. Less than 6% of raw data was smoothed to minimize the observed errors. The smoothed process consisted to calculate the new values from the average of six values data observed before and after the error (three on each side).

Video Data

A second set of data was recorded on a 1.45 m wide exclusive off-street bicycle facility located in Montréal, using a video collection unit. This recorded the times of passages of bicycles passing over a short section of 2 m. A camera mounted on a 6-m high pole was filming the flow of bicycles during the p.m. peak hour. From the 324 cyclists recorded, 253 pairs of cyclists were identified. These pairs were within platoons of two to 15 cyclists. The data was extracted manually with the Windows Movie Maker utility from Microsoft that allows the processing with a precision of one-thirtieth ($1/30$ th) of a second. We could then obtain the individual speed, the spacing, the gap, and the headway for each pair of cyclists. The dataset would then yield through mathematical transformations to flow rate, density and speed.

Global Data Analysis

The dataset collected provides a rich amount of information on the dynamics of the cyclist and his behavior. Regarding of the basic parameters (speed, acceleration, and spacing), [Table 1](#) gives the general statistics compiled using data collected with the GPS device as well as the recorded video.

These values are quite similar to those found in the literature. For example, the average speed used by the *Highway Capacity Manual 2010* (4) is 5.7 m/s. In general, the average speed reported in the literature (5, 6) varies between 4.7 and 6.9 m/s. As for the acceleration, one of the rare studies on the subject in Denver, Colorado, (6) reported an average acceleration of 0.24 m/s² and an average deceleration of -0.27 m/s² on an off-street facility.

TABLE 1 Statistics of Bicycles Dynamics

Parameters	Minimum	Maximum	Average	85th Percentile	Standard Deviation	<i>N</i>
Speed (m/s) (from video)	2.00	8.00	5.67	6.42	0.99	324
Acceleration (m/s ²) (from GPS)	—	1.34	0.20	0.40	0.21	340
Deceleration (m/s ²) (from GPS)	-1.81	—	-0.20	-0.03	0.22	340
Spacing (m) (from video)	2.89	62.91	14.78	24.93	10.09	253
Spacing (m) (from GPS)	2.22	61.10	13.2	23.17	10.00	340

BIKE-FOLLOWING MODEL AND RESULTS

As bicycle traffic increases, a cyclist increasingly needs to evaluate and adjust his relative position to other cyclists, thus the importance of modeling the bike-following situation. A bike-following model is a single mathematical equation combining direct and indirect physical and operational characteristics of the bicycle (length, speed, and acceleration), human factors (perception, decision, and action), and the cyclist's immediate environment (heavy, moderate, or free-flow traffic). Overall, the microscopic model is formulated using the following logic:

$$\text{Reaction}_{(t+\Delta t)} = f[\text{Sensibility}_{(t)}, \text{Stimulus}_{(t)}]$$

Over the past 50 years, close to 100 car-following models have been developed (7). These models were originally intended for the understanding the traffic dynamic, more specifically, the phenomenon of congestion. These models are also used in traffic safety analysis, and since the late 1960s, they are imbedded in microscopic traffic simulators. Currently, microscopic simulation is extensively used for planning, training, and research. As of the mid-1990s, microscopic simulation has become an integral part of intelligent transportation systems.

Microscopic Data Analysis

The data collected with the GPS device allowed reproducing the actual behaviour of both cyclists during the trip using one-second time steps. Figure 1 shows the trajectory of the pair of cyclists, their speed and acceleration profiles as well as the spacing maintained between the two bicycles.

Figure 1 shows the correlation between the each cyclist's dynamic behavior, especially when the spacing between them is reduced, which occurs around the 60th second. At this moment, the follower is forced to adjust his speed to the leading cyclist because the follower can not overtake the leader. The follower's set point was to start 10 s after the start of the first cyclist, then he had to catch up and follow the leader without overtaking him.

Thus, there exists a given spacing between two cyclists beyond which the follower moves freely and below which he needs to adjust his speed to the leader to maintain a safe distance while trying to maintain his desired speed. This can be better illustrated using a phase plane of

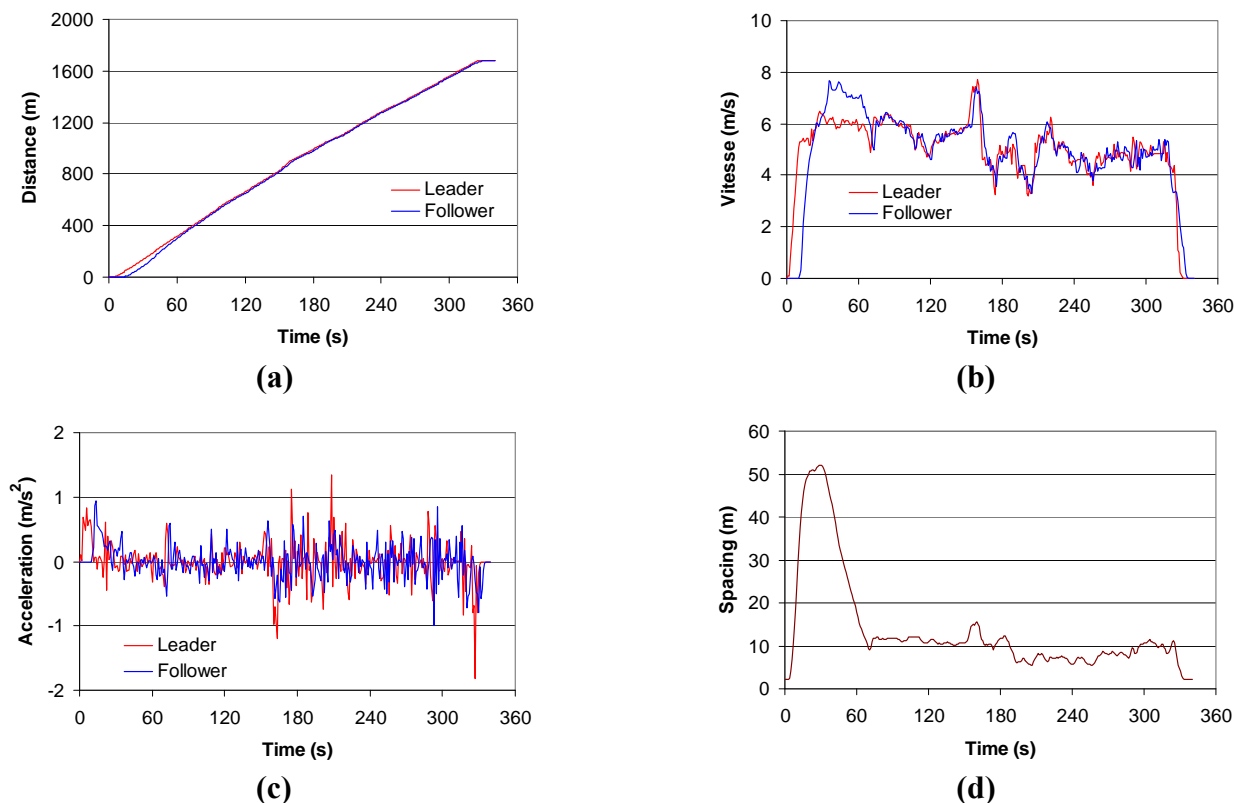


FIGURE 1 Behavior of a pair of cyclists in following situation: (a) time–space diagram; (b) speed profile; (c) acceleration profile; and (d) spacing profile.

the first order, also known as Action Point Model, as shown in Figure 2. This graph links the relative velocity (speed difference of the two cyclists) to the spacing between the two cyclists (8).

From Figure 2, it is found that the follower enters the leading cyclist's influence zone when their spacing is about 16 m. Kahn and Raksuntorn (6) found this interaction distance equal to 21 m. In a natural process, the follower tries to minimize the difference between his and the leader's speed while maintaining a safe and comfortable distance. The equilibrium is achieved through several adjustment cycles by the follower. This equilibrium point is not static but varies depending on the relative speed and spacing.

In order to adjust his speed, the follower must continually be aware of his position relatively to his equilibrium point. This information is provided through the distance perception threshold (DPT). This is defined as the minimum level of change in spacing that can be perceived by the cyclist. At this threshold, the cyclist becomes aware that he is moving away from his equilibrium point. If he is negatively moving away from this point, he will increase his speed and if he is positively moving away, he will reduce his speed. The simplest equation, developed by Evans (9), was chosen and adapted to bicycle traffic.

The same figure shows the minimal distance the cyclist keeps at all times to avoid colliding with the leading cyclist, when overtaking is not possible. This minimal spacing corresponds to the length of a bicycle (1.8 m) plus a minimal gap. According to field data collected with the GPS

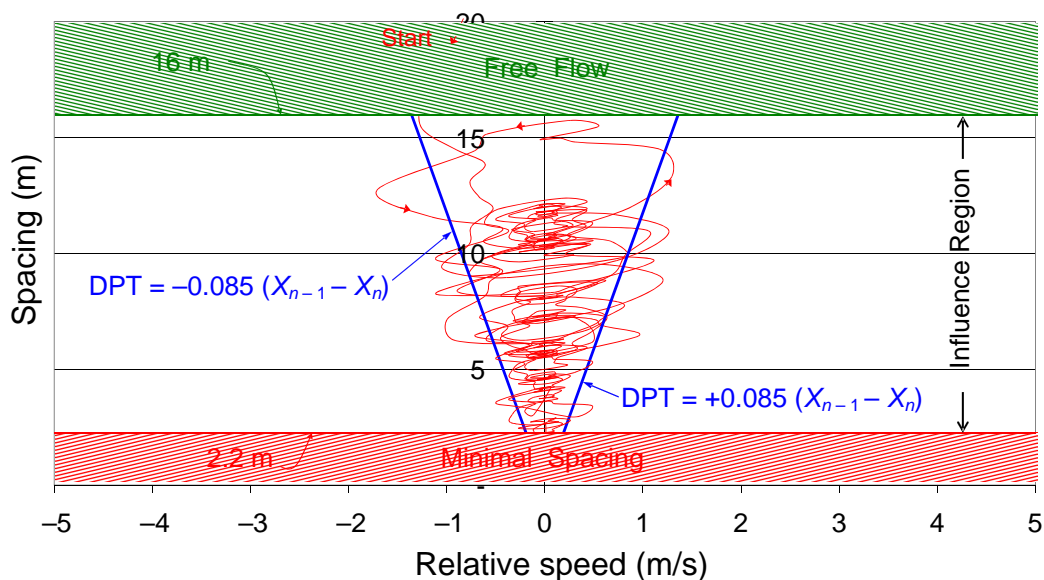


FIGURE 2 Phase plane trajectory for a pair of cyclists in following situation. (DPT = distance perception threshold.)

device, this minimal gap is about 0.42 m; therefore, resulting in a minimal spacing of 2.22 m. However, the video data indicated that some following cyclists had a lateral offset to the leader but just behind him in the same lane. So, there was no spacing between the two cyclists. In the most case, this situation was a prelude to the overtaking maneuver. These cases were not taken into consideration for this study.

Bike-Following Models

Five models representing the bike-following situation were investigated. The choice is mainly based on the popularity of models and their theoretical concept. In addition, a sixth model proposed by the authors has also been the object of the analysis. Table 2 shows the models used to reproduce the behaviour of the following cyclist according to his own behavior and the behavior of the leading cyclist. The first four models are derived from existing car-following models, the fifth developed specifically for bikes and the last is proposed by the authors.

The first three models simulate the acceleration and the other three simulate the speed. The model proposed by the authors incorporates two constraints. The first constraint reflects a minimum spacing (S_j) between two stopped cyclists. A second constraint limits the maximum speed of the cyclist, when the free-flow conditions prevail. It is assumed that this speed corresponds to the 85th percentile of observed speeds.

For the calibration process, the values of the proposed models parameters were optimized using an iterative function minimizing the sum of squared deviations. This optimization allowed reproducing as faithfully as possible the four variables describing the behaviour of two cyclists in a following situation. These variables are speed, acceleration, position and the spacing between the two bicycles for every time step (1 s).

TABLE 2 A List of Bike-Following Models Investigated

Models	Formula
Chandler (10)	$a_n(t+T) = \lambda \left[v_{n-1}(t) - v_n(t) \right]$
Gazis (11)	$a_n(t+T) = \alpha \frac{v_n(t)^m}{\left[x_{n-1}(t) - x_n(t) \right]^l} \left[v_{n-1}(t) - v_n(t) \right]$
Pipes (12)	$a_n(t+T) = CW \frac{v_{n-1}(t) - v_n(t)}{\left[x_{n-1}(t) - x_n(t) \right]^2}$
Gipps (13)	$v_n(t+T) = \text{Min} \left\{ \begin{array}{l} v_n(t) + 2,5aT(1-v_n(t)/V)\sqrt{0,025 + v_n(t)/V} ; \\ bT + \left[b^2T^2 - b \left\{ 2[x_{n-1}(t) - x_n(t) - s] - v_n(t)T \right. \right. \\ \left. \left. - v_{n-1}(t)^2/b^* \right\} \right]^{1/2} \end{array} \right\}$
Raksuntorn (6)	$v_n(t+\Delta t) = \alpha_1 v_n(t) + \alpha_2 \left[x_{n-1}(t) - x_n(t) \right] + \alpha_3 \left[v_{n-1}(t) - v_n(t) \right]$
Proposed Model (Manar)	$v_n(t+T) = \text{Min} \left\{ V_f ; \delta \left(x_{n-1}(t) - x_n(t) - S_j \right) \right\}$

In addition, for each variable, we calculated the regression between observed and theoretical values obtained by each model. The value of the slope of the regression and the multiple correlation coefficients (R^2) were obtained. Also, the Fisher significance test was calculated to assess the similarity between the observed and theoretical values. Table 3 presents the optimized parameters and the statistical comparison between the observed and theoretical data.

Overall, all models can reproduce fairly well the behaviour of the cyclist in a following situation. Statistically, all models are significant for reproducing speed, acceleration and position of the following cyclist. However, to estimate the spacing between the two cyclists, only the Raksuntorn’s model produced a statistically significant result.

Figure 3 shows the curves for speed (3a), acceleration (3b), and spacing (3c) as modeled by the six bike-following models. These simulated results are compared with the following and leading cyclists’ actual profiles obtained through the GPS device. For the case of spacing, it is clear that the models that simulate the speed rather than acceleration produce better results. But when the first 70 s during which the follower was trying to catch up the leader are removed (Figure 3d), all models perform well.

TABLE 3 Bike-Following Models Comparisons

Models/ Parameters		Chandler	Gazis	Pipes	Gipps	Raksuntorn	Manar
Optimized coefficients		$\lambda = 1.7$	$\alpha = 1.7$ $m = 0.01$ $l = 0.17$	CW = 680	$a = 2.0$ $T = 1.0$ $V = 10.0$ $s = 3.0$ $b = -2.0$ $b^* = -2.0$	$\alpha_1 = 0.96$ $\alpha_2 = 0.02$ $\alpha_3 = 0.35$	$V_f = 6.42$ $\delta = 0.85$ $S_f = 2.22$
Speed	Slope	0.999	1.008	1.006	0.994	1.000	0.995
	R^2	0.971	0.981	0.978	0.976	0.985	0.972
	F_{cal}/F_{crit}	0.860	0.934	0.909	0.866	0.929	0.935
Position	Slope	1.010	1.012	1.011	1.000	1.000	0.999
	R^2	0.999	0.999	0.999	0.999	0.999	0.999
	F_{cal}/F_{crit}	0.800	0.800	0.801	0.794	0.790	0.796
Acceleration	Slope	0.346	0.914	0.646	0.752	1.387	1.105
	R^2	0.935	0.957	0.936	0.915	0.832	0.930
	F_{cal}/F_{crit}	0.111	0.637	0.326	0.729	0.932	0.515
Spacing	Slope	0.656	0.585	0.605	1.315	1.356	0.534
	R^2	0.594	0.599	0.588	0.703	0.838	0.860
	F_{cal}/F_{crit}	7.752	3.089	8.094	1.467	0.849	1.307

NOTE: The ratio Computed Fisher/Critical Fisher ≤ 1 means that one accepts the hypothesis that variances of observed and theoretical values are not statistically different for a level of confidence of 99% ($\alpha = 0.01$; $D_f = 340$).

Microsimulation with VISSIM

Moreover, the data collected were used to compare with results produced by micro-simulation software VISSIM (14). During the simulation with a pair of bikes, the leader’s positions and speeds were adjusted for each second according to collected data, and the software determine the speeds and acceleration/deceleration of the follower. The experiment was done with default parameters offered with software and with calibrated acceleration–deceleration and driving behaviour parameters. As illustrated in Figure 4, the speed of followers produced with default parameters is very sensitive and unstable, but after calibration effort, the result was relatively smooth and convincing. The sum of squared deviation between the observed follower’s speeds was reduced from 1,051 (with default parameters) to 478 (after calibration).

BICYCLE TRAFFIC FLOW MODELS

The *Highway Capacity Manual 2010* (4) recommends a capacity of 2,000 cyclists per hour per lane. However, the capacity values reported in the literature (5) vary from 1,500 to 5,000 bicycles per hour depending on bicycle facility width. Because traffic flow on bicycle facilities is usually

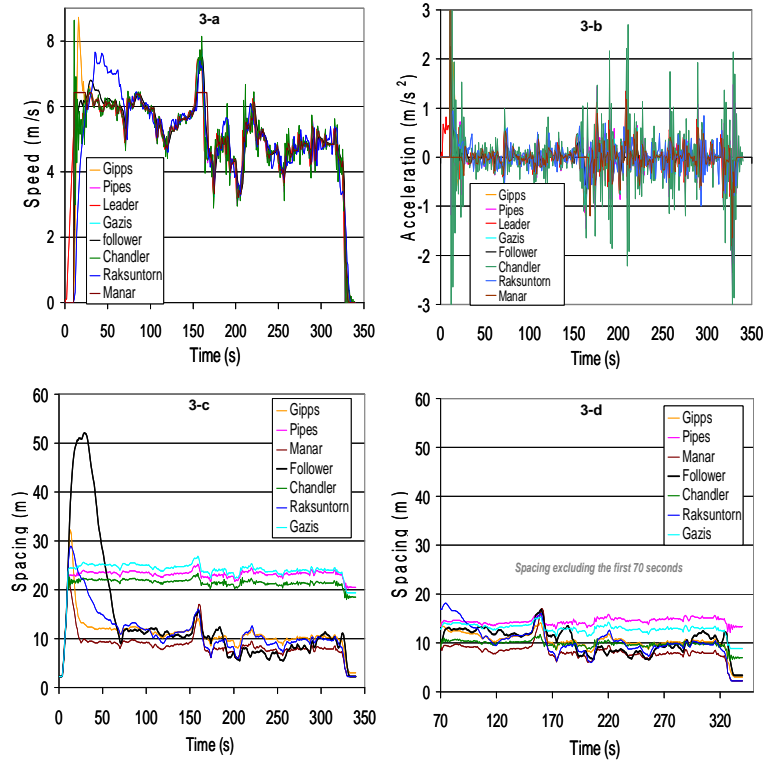


FIGURE 3 Bike-following models results: (a) speed, (b) acceleration, (c) spacing, and (d) period during which the follower trying to catch up with the leader is removed.

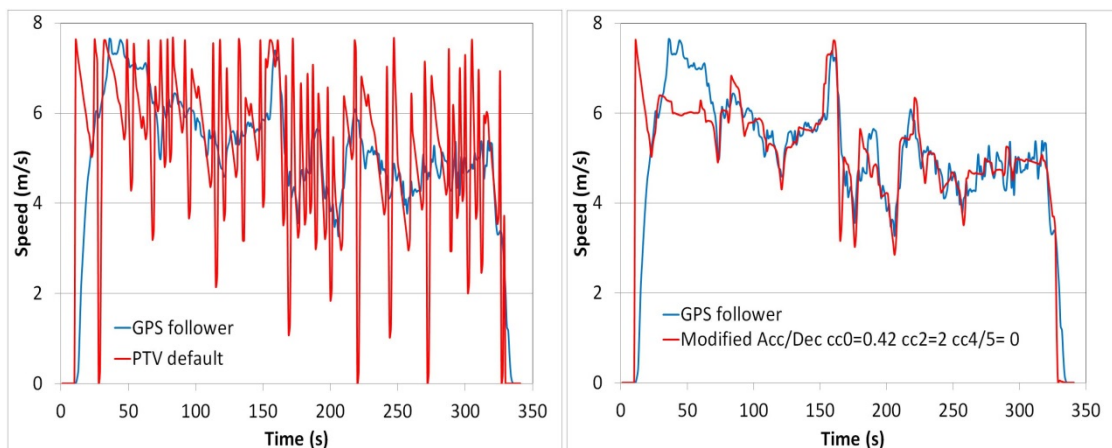


FIGURE 4 VISSIM simulation results.

low, capacity can rarely be observed.

The data collected on the off-street facility with video recorder were used to define the fundamental relationships of bicycle traffic. Thus, the flow was calculated from the inverse of headways, and the density from the inverse of spacing.

In order to construct fundamental relationships curves, it was required to determine the theoretical maximum density of the facility. The theoretical maximum density from the

experimentation was 0.33 bicycles/m², which is equivalent to 475 bicycles/km for a 1.45 m wide bicycle lane. AASHTO (15) recommends nearly the same value with 0.32 bicycles/m². These values are based on minimal manoeuvring space required for a cyclist.

Four existing models were investigated for fit. Table 4 presents the models chosen, the optimized coefficients of their parameters and the R² values.

These models were calibrated and the resulting curves are presented in Figure 5. For the flow–density relationship, all four models reproduced an acceptable fit with observed data. However, the speed–density correlation of all four models with field data is relatively low. This is explained by the high dispersion of speed when bicycle traffic is fluid. This is actually the reflection of differences in physical abilities of cyclists in this sample. In all cases, the shapes of the empirical results are fully consistent with the modeled results.

COMPARISON BETWEEN BIKE–CYCLIST AND CAR–DRIVER SYSTEMS

To compare the two systems, the same two cyclists drove two cars in a following situation. The same GPS devices and set of rules were used for the bicycle-following experiment as for the car-following test.

Despite differences in the values of driving parameters for the two systems, the two phase plans (Figures 2 and 6) show a similarity in the driving behavior.

TABLE 4 Fundamental Relationships Models Investigated

Models	Formula	Optimized Coefficients and Field Limit Values	R ² of Observed and Modeled Data Values	
			Flow	Speed
Newell (16)	$v = v_f \left[1 - e^{-\frac{\lambda}{v_f} (1/k - 1/k_j)} \right]$	$v_f = 24$ $k_j = 475$ $\lambda = 13,200$	0.87	0.36
Pipes-Mujal (12)	$v = v_f \left[1 - (k/k_j)^n \right]$	$v_f = 24$ $k_j = 475$ $n = 1.65$	0.84	0.35
Northwestern (modified ^a) (17)	$v = v_f e^{-1/2 (k/k_o)^\alpha}$	$v_f = 24$ $k_o = 250$ $\alpha = 3.4$	0.74	0.31
Van Aerde (18)	$k = \frac{1}{c_1 + \frac{c_2}{v_f - v} + c_3 v}$	$v_f = 24$ $c_1 = 0.00187135$ $c_2 = 0.00561404$ $c_3 = 0.00010046$	0.83	0.34

^a The constant value of α in the original model was 2.

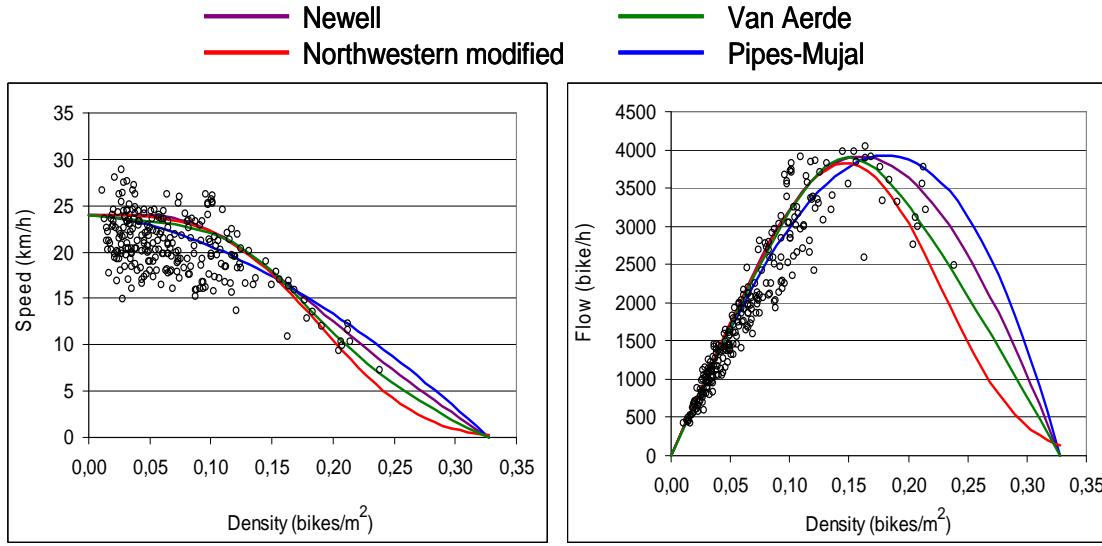


FIGURE 5 Fundamental relationships comparison with experimental data.

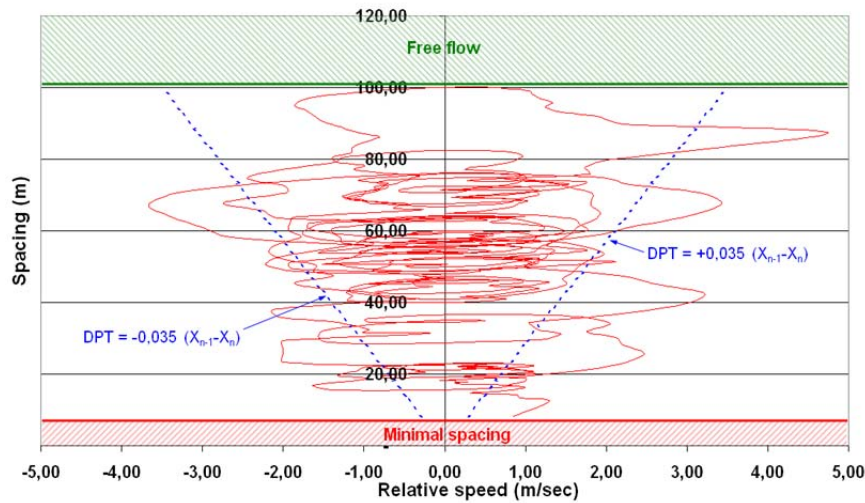


FIGURE 6 Phase plane trajectory for a pair of cars in following situation.

However, to obtain an objective comparison, a new concept is introduced: a Normalized Phase Plan (Figure 7). The two phase plans are normalized according to the maximum values of each system. Thus, by dividing the values of relative speed for each system by its maximum value, and dividing the values of spacing for each system by its maximum value, the two phase plans obtained can be directly compared.

Figure 7 clearly shows that following drivers, indifferently of the mode used, adopt the same driving logic when seeking the acceptable equilibrium between their desired speed and maintaining a safe distance from the leader. This figure also shows that the amplitude of most phase cycles fall within a 40% deviation of relative speed. It means that, whatever the mode used, the following driver accelerates or decelerates at the same amplitude but not necessarily at the same rate.

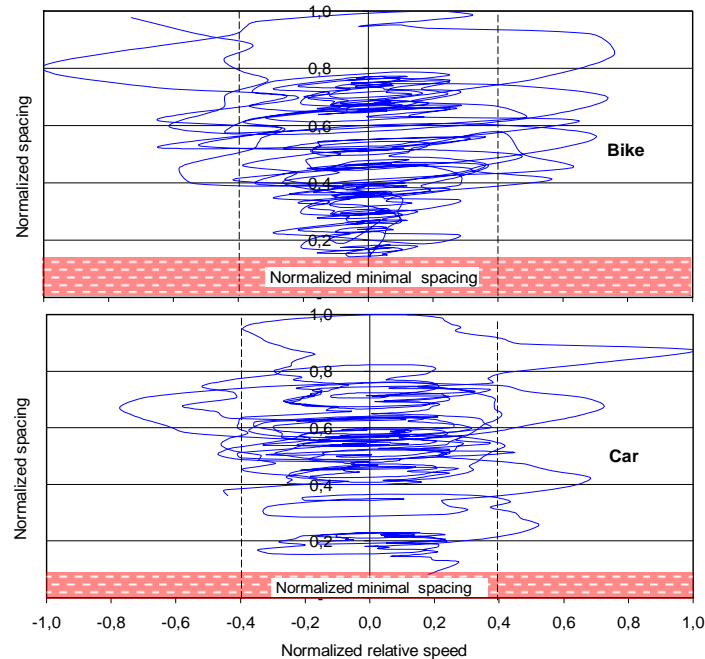


FIGURE 7 Normalized phase plane trajectories for a pair of cars and bicycles.

FINDINGS AND CONCLUSIONS

With the increasing popularity of cycling in Montréal, we now see an increased intensity of traffic on bicycle facilities. The authors take advantage of the advances in data gathering technologies to automatically collect data and investigate fundamental as well as bike-following models. In the latter case, a new bike-following model is proposed.

Two types of data were recorded as part of this research. First, a pair of cyclists in a following situation was tracked using a GPS device. Second, a video collection unit was used to film a flow of cyclists on a busy bicycle facility at a fixed spot location. The findings can be summarized as follow:

- There is a distance between two cyclists beyond which the follower moves freely and below which he needs to adjust his speed to the leader to maintain a safe distance when the overtaking is not possible. This zone of influence starts when the spacing is about 16 m. The minimal spacing, S_j , is 2.2 m including the bicycle length (1.8 m).
- Using the instantaneous flow rate and measuring the space needed by cyclists to manoeuvre, it was found that flow can reach nearly 2700 bicycles/h/m and the maximum density (K_j) can reach 0.33 bicycle/m² for this 1.45 m width bicycle facility.
- Investigating the fundamental relationships, it was found that the experimental data are fully consistent with the results obtained with theoretical relationships. However, correlation for Speed-Density relationship is relatively low due to greater speed dispersion reflecting the large differences in physical abilities of cyclists. On the other hand, flow–density relationships offer a better fit and Newell’s model performs better than the others.
- Six bike-following models were calibrated. All models perform well to estimate

speed and acceleration, but only Raksuntorn's model is statistically significant for estimating spacing. However, when the first 70 s of data are removed, all models estimate well the spacing. These first 70 s corresponds to the time the follower took to catch up to the leader. The acceleration values during this time were quite high.

- Using the same experimental conditions, bike-following and car-following observations were compared using normalized phase plans. It can be concluded that a given pair of drivers behave similarly independently of the type of vehicle used.
- Simulation software like VISSIM (14) can be used to simulate bike interaction, but the calibration of parameters is needed to adapt the software to the characteristics of cyclist.

These bike-following models can be use in commercial simulation packages like VISSIM (14), CORSIM (19), or SIMTRAFFIC (20) to account for longitudinal bicycle dynamics. With additional research on cyclist behaviour, and their interaction with other vehicles, the existing simulation tools or the next generation of modelling tools will be able to simulate real world mixed traffic conditions.

REFERENCES

1. Taylor, D., and W. J. Davis. Review of Basic Research in Bicycle Traffic Science, Traffic Operations, and Facility Design. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1674*, TRB, National Research Council, Washington, D.C., 1999, pp. 102–110.
2. Gould, G., and A. Karner. Modeling Bicycle Facility Operation: A Cellular Automaton Approach. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2140*, Transportation Research Board of the National Academies, Washington, D.C., 2009, pp. 157–164.
3. Faghri, A., and E. Egyhaziova. Development of Computer Simulation Model of Mixed Motor Vehicle and Bicycle Traffic on an Urban Road Network. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1674*, TRB, National Research Council, Washington D.C., 1999, pp. 86–93.
4. *Highway Capacity Manual 2010*. Transportation Research Board of the National Academies, Washington, D.C., 2010.
5. Allen, D. P., N. Roupail, J. E. Hummer, and J. S. Milazzo II. Operational Analysis of Uninterrupted Bicycle Facilities. In *Transportation Research Record 1636*, TRB, National Research Council, Washington, D.C., 1998, pp. 29–36.
6. Raksuntorn, W., and S. I. Khan. Behavior of Bicyclists in Following. Presented at the 85th Annual Meeting of the Transportation Research Board, Washington D.C., 2006.
7. Brackstone, M., and M. McDonald. Car-Following: A Historical Review. *Transportation Research Part F*, Vol. 2, No. 4, 1999, pp. 181–196.
8. Todosiev, E. P. The Action Point Model of the Driver Vehicle System. Report No. 202A-3. Ohio State University, Engineering Experiment Station, Columbus, Ohio, 1963.
9. Evans, L., and R. Rothery. Detection of the Sign of Relative Motion When Following a Vehicle. *Human Factors*, Vol. 16, No 2, 1974, pp. 161–173.
10. Chandler, R. E., R. Herman, and E. W. Montroll. Traffic Dynamics: Studies in Car Following. *Operations Research*, Vol. 6, No. 2, 1958, pp. 165–184.
11. Gazis, D. C., R. Herman, and R. W. Rothery. Nonlinear Follow the Leader Models of Traffic Flow. *Operations Research*, Vol. 9, No. 4, 1961, pp. 545–567.
12. Pipes, L. A. Car-Following Models and the Fundamental Diagram of Road Traffic. *Transportation Research*, Vol. 1, 1967, pp. 21–29.

13. Gipps, P. G. A Behavioral Car Following Model for Computer Simulation. *Transportation Research B*, Vol. 15, 1981, pp. 105–111.
14. PTVAMERICA. VISSIM version 6.00-15. Portland, Oregon, 2014.
15. AASHTO. A Policy on Geometric Design of Highways and Streets, 4th edition, 2001.
16. Newell, G. F. Nonlinear Effects in the Dynamics of Car Following. *Operation Research*, Vol. 9, 1961, pp. 209–229.
17. Drake, J. L. S. J. S., and A. D. May. A Statistical Analysis of Speed–Density Hypotheses. *Highway Research Record 156*, HRB, National Research Council, Washington, D.C., 1967, pp. 53–87.
18. Van Aerde, M. Single Regime Speed–Flow–Density Relationship for Congested and Uncongested Highways. Presented at 74th Annual Meeting of the Transportation Research Board, Washington, D.C., 1995.
19. Federal Highway Administration. CORSIM User Manual. FHWA, McLean, Va., 1996.
20. Trafficware. SIMTRAFFIC 8. Sugar Land, Tex., 2011.

POSTER PRESENTATION

Gaussian Approximation for Modeling Traffic Flow on a Homogeneous Road Segment

BRIAN THOMAS VACHTA

XIAO QIN

JUNG-HAN KIMN

South Dakota State University

A Gaussian approximation of the stochastic model by Jabari and Liu proposed is implemented for a case study. The Gaussian approximation estimates the mean traffic density from the long-term trend of a stochastic model, and a variance calculated from the deviance of the stochastic model and its long-term limit. The mean is the deterministic Godunov scheme. Through this case study the model is shown capable of producing accurate queue size estimates in a viable amount of time for cycle-by-cycle traffic estimation on an arterial road. Model variance is controllable by choice of time step; however, smaller time steps limit the ability of the model to produce a range of solutions which contain the ground truth observation.

INTRODUCTION

Traffic congestion on roads is a continuing concern around the world. In the past this has been addressed by adding capacity to infrastructure, i.e., adding more lanes to highways and adding more roads. This solution is largely subject to budget constraints.

Another solution that has been adopted more recently is operating existing infrastructure more efficiently. This strategy involves using knowledge of real-time traffic conditions to reduce congestion (*I*). Examples of this include ramp metering and advanced wait time notice on freeways. In the context of traffic lights, real-time traffic data allows for more efficient traffic signal timings.

Traffic signals historically operate by fixed-cycle lengths of red time and green time. Recently, many fixed-cycle lengths require actuation in order for certain phases of the traffic cycle to occur, i.e., left-turn signals that are only included in the cycle if a vehicle is detected in the left-turn lane. Actuated signals can be more efficient than fixed-cycle timings by skipping unnecessary phases of a signal cycle, shortening the time vehicles wait at the signal. However, traffic has to stop at the stop line in order to actuate the signal. Adaptive (predictive) traffic signal timings, which are based on real-time arrival rates of traffic, further improve efficiency by optimizing signal cycle lengths according to expected queue sizes (number of vehicles stopped behind stop line). The expected queue sizes identify where the bulk of traffic is located. As an example, if the location of the bulk of traffic is located, that traffic could progress through continual green lights.

In the development of cost-efficient methods to predict traffic conditions, point-source detectors such as inductance loop detectors are cheaper to implement; however, they lack the accuracy that video cameras have. This has led to the development of traffic flow models to predict traffic conditions based off of inductance loop detector data.

TRAFFIC MODELING OVERVIEW

Traffic flow theory models estimate traffic evolution for various levels of vehicle aggregation. The most common types of models are microscopic, mesoscopic, and macroscopic.

Microscopic traffic flow models have high detail and depend on driver's behavior interaction between individual vehicles. These behaviors include driver's choice of time distance between other vehicles, choice of speed, acceleration, deceleration, and lane changes. Microscopic models can simulate complex traffic conditions. In order to accurately model traffic, a large amount of parameters are required. A disadvantage to using microscopic models to predict traffic conditions is that real-time point-source data is usually aggregated over 1- to 5-min intervals, thus model updates are not as precise when updated with aggregated data.

Mesoscopic models have medium detail, describing the behavior of small groups of vehicles, segregated by speed and position.

Macroscopic models have low level of detail as they represent aggregated levels of flow without considering individual vehicle behavior. These models are able to produce results fast and are able to use aggregated real-time data without loss of accuracy, which motivates their application.

Lighthill and Whitham and Richards proposed a first-order deterministic model which is widely used throughout macroscopic modeling—the Lighthill-Whitham and Richards (LWR) model (2, 3). The LWR model has been used to describe queue buildup and discharge in traffic. It shows the existence of shock waves in traffic; however, its solutions assume that traffic can adjust its speed instantaneously. It lacks diffusive terms which would model drivers' reaction to changes in traffic such as shock waves.

While deterministic models have been able to capture important physical phenomenon, these models are typically dependent on an assumed flow–density relationship. Flow–density relationships admit large fluctuations in behavior (4, 5). This high variability is due to driver decisions, adverse weather conditions, etc., and has led to the development of stochastic microscopic and macroscopic models to capture this variation. A challenge in stochastic modeling is obtaining reasonable computation times. In an attempt to minimize this issue, many models have been created by adding Gaussian noise to deterministic models. Since Gaussian noise can be characterized by its mean and variance, this becomes less of a computational hassle. However the addition of Gaussian noise could contradict physical phenomenon by producing negative traffic densities as well as avoiding mean dynamics of the original deterministic model as the dynamic equations are nonlinear (6).

A computationally viable model is proposed in Jabari and Liu (7) which uses the long-term trend of the stochastic Markovian model as the mean of the model (6). The covariance of the mean is derived from the deviance of the stochastic model and its deterministic long-term limit. The long-term trend or fluid limit of the stochastic model is the deterministic Godunov scheme, which is a method for solving the macroscopic LWR model. The variance is controllable by the time step such that nonnegative densities ranges can be obtained for sufficiently small time step.

Because of the linearity of the stochastic model, the state can be compared to real data using the linear Kalman filter to improve mean traffic density estimations. Using the linear Kalman filter is faster than using nonlinear filters and is a unique feature of this model.

OBJECTIVE

The primary objective of this paper is to investigate the use of the Gaussian approximation proposed in Jabari and Liu (6). This method is chosen for its ability to guarantee of nonnegative traffic flow and that the variance is derived from deterministic model parameters such that the model is computationally feasible to implement in traffic control systems. We implement the model for a case study in which we wish to describe traffic conditions downstream on a homogeneous road with no lane changes from knowledge of upstream conditions.

LITERARY REVIEW

Lighthill, Whitham, and Richards Model

Macroscopic models describe traffic evolution as aggregated amounts of traffic. These models use traffic density, flow rates, and a fundamental relationship to characterize traffic. Lighthill and Whitham and Richards proposed the following first-order macroscopic model from hydrodynamic theory (4, 5):

$$\rho_t + q_x = 0 \quad (1)$$

for traffic density $\rho(x,t)$, and flow $q(x,t)$. This partial differential equation (PDE) represents the change of vehicles in a road segment as the difference of the flow of vehicles in and out of the road. This was proposed as a model of traffic flow with the assumption that the flow of traffic is a function of density: $q(x,t) = Q[\rho(x,t)]$. This reduces the PDE to an ordinary differential equation:

$$\frac{\partial \rho(x,t)}{\partial t} + Q'(\rho) \frac{\partial \rho(x,t)}{\partial t} = 0 \quad (2)$$

The general solutions to Equation 2 are a family of straight lines in the $x-t$ plane. Each solution has wave speed, or slope, $Q'(\rho)$ characterized by the value of ρ . A popular method to model traffic with the LWR model is to divide a road into segments called cells, choose a flux function, and use initial conditions to solve 2. By dividing the road into cells, the assumption of differentiability used to derive 2 does not hold, and only weak solutions are allowed. The weak solutions are not unique, such that various methods have been developed to physically model traffic flow. These methods are based on how shock waves and expansion waves are demonstrated.

Daganzo (8, 9) proposed piecewise linear sending and receiving functions in Equations 3 and 4 as flux functions to describe flow between cells as well as shock waves in traffic. These functions are used to create a flux function as seen in Figure 1. This method is a cell transmission model (CTM), which is a special case of Godunov's scheme (10). Daganzo's CTM was developed to solve the LWR model for complicated network topologies.

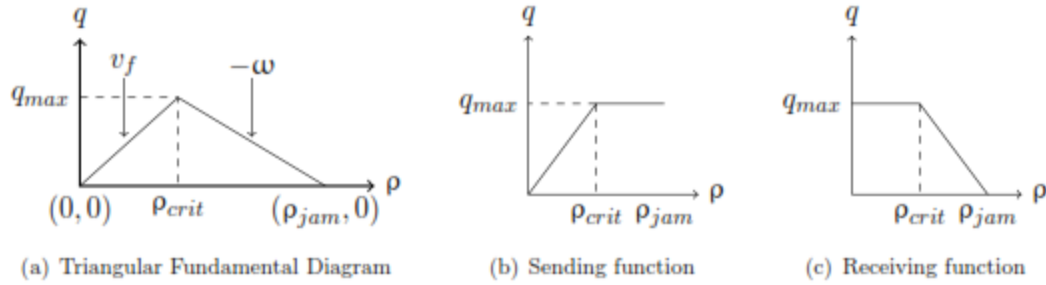


FIGURE 1 Example of Daganzo's flux function (a) composed from the minimum of (b) a sending function and (c) a receiving function.

$$S_e(\rho(i, t)) = \min\{v_f \rho(i, t), q_{\max}\} \quad (3)$$

$$R_e(\rho(i, t)) = \min\{q_{\max}, \omega(\rho_{\text{jam}} - \rho(i, t))\} \quad (4)$$

The subscript 'e' denotes an equilibrium or empirically derived (fundamental) flow–density relationship.

For ease of notation, we declare $y(i + .5, t)$ to be the vector of densities in cells adjacent to the boundary at $i + .5$ at time t for all $i \in C$.

$$y(i + .5, t) = \begin{bmatrix} \rho(i, t) \\ \rho(i + 1, t) \end{bmatrix} \quad (5)$$

This notation is used in calculation of CTM flux functions, which calculate shock waves at the boundaries of cells. $|C|$ denotes the cardinality, or size of the set of cells in C . This is used to denote the index of the last cell. The procedure for Daganzo's CTM is found in [Algorithm 1](#).

Algorithm 1 CTM

Assume initial density $\rho(i, 0)$ for all cells $i \in C$.
 Initialize model with fundamental parameters.
 Compute flux function at boundaries for each cell $i = 1$ to $|C|$ from $k = 1$ to end:

$$\begin{aligned} \lambda(y(.5, t_k)) &= R_e(\rho(1, t_k)) \\ \lambda(y(|C| + .5, t_k)) &= S_e(\rho(|C|, t_k)) \\ \lambda(y(i + .5, t_k)) &= \min\{S_e(\rho(i, t_k), R_e(\rho(i + 1, t_k))\} \end{aligned}$$

Update new densities:

$$\rho(i, t_{k+1}) = \rho(i, t_k) + \frac{\Delta t}{\ell_i} (\lambda(y(i - .5, t_k)) - \lambda(y(i + .5, t_k)))$$

Next k

Gaussian Approximation

The Gaussian approximation model uses the CTM as an approximation to the long-term trend of the stochastic model in Jabari and Liu (6). The stochastic model is a microscopic queueing model which counts vehicle departures from cells based on stochastic time headways. The deviance between the stochastic and its fluid limit, outlined in Jabari and Liu (7), allows us to calculate the covariance of the deterministic model based on the stochastic model's probabilistic deviance from it. This allows a range of solutions to be given for density estimates based on a stochastically derived model. The deviance is derived as a stochastic differential equation, with solution in differential form given in Equation 6 for initial covariance matrix $\Psi(0)$, where $D(t)$, B , and $\Gamma(t)$ are given in Equations 9, 10, and 11. Equation 6 is approximated using Riemann sum integration:

$$\frac{d\Psi(t)}{dt} = D(t)\Psi(t) + \Psi(t)D(t)^T + B\Gamma(t)\Gamma(t)^TB^T \quad (6)$$

$$\int_t^{t+\Delta t} \frac{d\Psi(t)}{dt} dt = \Psi(t + \Delta t) - \Psi(t) \quad (7)$$

$$\cong D(t)\Psi(t)\Delta t + \Psi(t)D(t)^T\Delta t + B\Gamma(t)\Gamma(t)^TB^T\Delta t$$

Such that:

$$\Psi(t + \Delta t) \cong \Psi(t) + D(t)\Psi(t)\Delta t + \Psi(t)D(t)^T\Delta t + B\Gamma(t)\Gamma(t)^TB^T\Delta t \quad (8)$$

This solution serves as the covariance of the data in the cells. The diagonal elements of the covariance matrix are the individual variances of each corresponding cell's density (Figure 2).

$D(t)$ is defined to be a $|C| \times |C|$ matrix with row elements given by Equation 9 for each cell $i \in C$.

The middle element is the diagonal element of row i ,

$$\frac{1}{\ell_i} \left[\dots 0 \frac{\partial \lambda(\bar{y}(i - .5, t))}{\partial \bar{\rho}(i - 1, t)} \frac{\partial \lambda(\bar{y}(i - .5, t))}{\partial \bar{\rho}(i, t)} - \frac{\partial \lambda(\bar{y}(i + .5, t))}{\partial \bar{\rho}(i, t)} - \frac{\partial \lambda(\bar{y}(i + .5, t))}{\partial \bar{\rho}(i + 1, t)} 0 \dots \right] \quad (9)$$

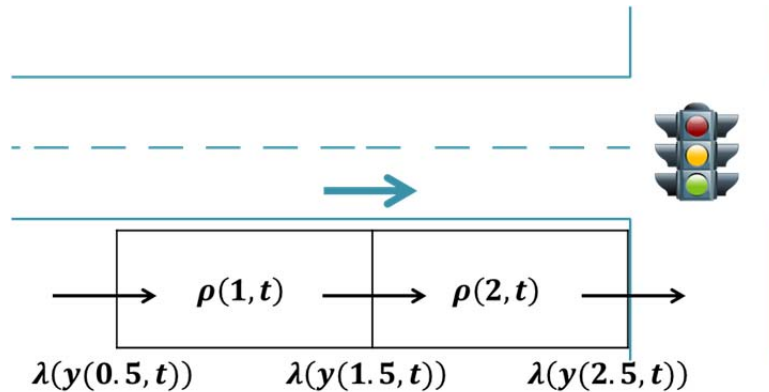


FIGURE 2 CTM with two cells.

\mathbf{B} is a $|C| \times |C| + 1$ matrix used to correctly combine flux terms in $\Gamma(t)$ with the corresponding cell length,

$$\mathbf{B} \equiv \begin{bmatrix} \frac{1}{\ell_1} & -\frac{1}{\ell_1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\ell_2} & -\frac{1}{\ell_2} & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \frac{1}{\ell_{|C|}} & -\frac{1}{\ell_{|C|}} \end{bmatrix} \quad (10)$$

$\Gamma(t)$ is a $|C| + 1 \times |C| + 1$ matrix of Ito integrands,

$$\Gamma(t) \equiv \begin{bmatrix} \bar{c}\sqrt{\lambda(\bar{\mathbf{y}}(.5, t))} & 0 & \cdots & 0 \\ 0 & \bar{c}\sqrt{\lambda(\bar{\mathbf{y}}(1.5, t))} & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \bar{c}\sqrt{\lambda(\bar{\mathbf{y}}(|C| + .5, t))} \end{bmatrix} \quad (11)$$

Derivatives of Flux Functions

The matrix $\mathbf{D}(t)$ requires the computation of partial derivatives of flux functions, which are non-differentiable in the classical sense. We can rewrite flux equation in Algorithm 1 as the average of the sending and receiving functions minus the distance (norm) between the two:

$$\min\{S_e(\bar{\rho}_i), R_e(\bar{\rho}_{i+1})\} = \frac{1}{2}(S_e(\bar{\rho}_i) + R_e(\bar{\rho}_{i+1}) - |S_e(\bar{\rho}_i) - R_e(\bar{\rho}_{i+1})|) \quad (12)$$

$$= \frac{1}{2}(S_e(\bar{\rho}_i) + R_e(\bar{\rho}_{i+1}) - \sqrt{(S_e(\bar{\rho}_i) - R_e(\bar{\rho}_{i+1}))^2}) \quad (13)$$

for $\bar{\rho}_i, \bar{\rho}_{i+1} \in [0, \rho_{\text{jam}}]$. This allows the partial derivatives to be written:

$$\frac{\partial \lambda(\bar{\mathbf{y}}(i + .5))}{\partial \rho_i} = \frac{1}{2} \frac{dS_e(\bar{\rho}_i)}{d\rho_i} \left(1 - \frac{S_e(\bar{\rho}_i) - R_e(\bar{\rho}_{i+1})}{|S_e(\bar{\rho}_i) - R_e(\bar{\rho}_{i+1})|} \right) \quad (14)$$

$$\frac{\partial \lambda(\bar{\mathbf{y}}(i + .5))}{\partial \rho_{i+1}} = \frac{1}{2} \frac{dR_e(\bar{\rho}_{i+1})}{d\rho_{i+1}} \left(1 + \frac{S_e(\bar{\rho}_i) - R_e(\bar{\rho}_{i+1})}{|S_e(\bar{\rho}_i) - R_e(\bar{\rho}_{i+1})|} \right) \quad (15)$$

The partial derivatives exist everywhere except for $S_e(\bar{\rho}_i) = R_e(\bar{\rho}_{i+1})$. For this case, Equation 12 reduces to:

$$\min\{S_e(\bar{\rho}_i), R_e(\bar{\rho}_{i+1})\} = \frac{1}{2}(S_e(\bar{\rho}_i) + R_e(\bar{\rho}_{i+1})), \quad (16)$$

such that the partial derivatives are:

$$\frac{\partial \lambda(\bar{\mathbf{y}}(i + .5))}{\partial \rho_i} = \frac{1}{2} \frac{dS_e(\bar{\rho}_i)}{d\rho_i} \quad (17)$$

$$\frac{\partial \lambda(\bar{\mathbf{y}}(i + .5))}{\partial \rho_{i+1}} = \frac{1}{2} \frac{dR_e(\bar{\rho}_{i+1})}{d\rho_{i+1}} \quad (18)$$

A summary of the weak partial derivatives of the flux function are found in [Table 1](#).

TABLE 1 Flux Derivatives

	$S_e(\bar{\rho}_i) < R_e(\bar{\rho}_{i+1})$	$S_e(\bar{\rho}_i) = R_e(\bar{\rho}_{i+1})$	$S_e(\bar{\rho}_i) > R_e(\bar{\rho}_{i+1})$
$\frac{\partial \lambda(\bar{\mathbf{y}}(i + .5))}{\partial \rho_i}$	$\frac{dS_e(\bar{\rho}_i)}{d\rho_i}$	$\frac{1}{2} \frac{dS_e(\bar{\rho}_i)}{d\rho_i}$	0
$\frac{\partial \lambda(\bar{\mathbf{y}}(i + .5))}{\partial \rho_{i+1}}$	0	$\frac{1}{2} \frac{dR_e(\bar{\rho}_{i+1})}{d\rho_{i+1}}$	$\frac{dR_e(\bar{\rho}_{i+1})}{d\rho_{i+1}}$

Gaussian Approximation Algorithm

Algorithm 2 Gaussian Approximation

Initialize model with fundamental parameters and \mathbf{B} :

$$\begin{aligned} \bar{\rho}(\cdot, 0) &= 0 \\ \lambda(\bar{\mathbf{y}}(\cdot, 0)) &= 0 \end{aligned}$$

For $k = 1$ to end:

Compute flux function at boundaries for $i = 1$ to $|C| - 1$:

$$\begin{aligned} \lambda(\bar{\mathbf{y}}(.5, t_k)) &= R_e(\bar{\rho}(1, t_k)) \\ \lambda(\bar{\mathbf{y}}(|C| + .5, t_k)) &= S_e(\bar{\rho}(|C|, t_k)) \\ \lambda(\bar{\mathbf{y}}(i + .5, t_k)) &= \min\{S_e(\bar{\rho}(i, t_k), R_e(\bar{\rho}(i + 1, t_k))\} \end{aligned}$$

Update $\mathbf{\Gamma}(t)$, $\mathbf{D}(t)$.

Compute covariance:

$$\mathbf{\Psi}(t_{k+1}) = \mathbf{\Psi}(t_k) + \mathbf{D}(t_k)\mathbf{\Psi}(t_k)\Delta t + \mathbf{\Psi}(t_k)\mathbf{D}(t_k)^T\Delta t + \mathbf{B}\mathbf{\Gamma}(t_k)\mathbf{\Gamma}(t_k)^T\mathbf{B}^T\Delta t$$

Update mean densities for $i = 1$ to $|C|$:

$$\bar{\rho}(i, t_{k+1}) = \bar{\rho}(i, t_k) + \frac{\Delta t}{\ell_i} (\lambda(\bar{\mathbf{y}}(i - .5, t_k)) - \lambda(\bar{\mathbf{y}}(i + .5, t_k)))$$

Calculate confidence interval from variance and mean:

$$\begin{aligned} ci^+ &= \bar{\rho}(i, t_{k+1}) + 1.96\sqrt{\mathbf{\Psi}_{ii}(t_{k+1})\Delta t} \\ ci^- &= \bar{\rho}(i, t_{k+1}) - 1.96\sqrt{\mathbf{\Psi}_{ii}(t_{k+1})\Delta t} \end{aligned}$$

Nonnegative Densities

The diagonal elements of the covariance matrix are the variances for each cell density. The variance is used to create a confidence interval about the mean. Mean density values are automatically guaranteed to be non-negative by the calculation of flux between cells. The range of potential density values around the mean is not guaranteed to be nonnegative.

We know from (Jabari and Liu, 2013) that the covariance matrix converges to $\mathbf{B}\mathbf{\Gamma}\mathbf{\Gamma}^T\mathbf{B}^T dt$ for which, the diagonal elements are $\Psi_{ii} = \left(\frac{\bar{c}}{\ell_i}\right) (\lambda(\bar{y}(i-1, t_k)) - \lambda(\bar{y}(i, t_k))) dt$. The variance is then calculated by:

$$\begin{aligned} \int_t^{t+\Delta t} \left(\frac{\bar{c}}{\ell_i}\right) (\lambda(\bar{y}(i-1, u)) + \lambda(\bar{y}(i, u))) du &= \int_t^{t+\Delta t} E|\tilde{\rho}(i, u) - E\tilde{\rho}(i, u)|^2 du \\ &= \left(\frac{\bar{c}}{\ell_i}\right)^2 (\lambda(\bar{y}(i-1, t)) + \lambda(\bar{y}(i, t))) \Delta t \\ &= \Psi_{ii} \Delta t \end{aligned} \quad (19)$$

To control the range of solutions to be guaranteed non-negative with 95% confidence, one would choose the smallest of the time steps that satisfies the CFL condition (11) and $\bar{\rho}(i, t_{k+1}) - 1.96\sqrt{\Psi_{ii}(t_{k+1})\Delta t} > 0$. For the latter part of the condition we have:

$$\Delta t \leq \frac{\bar{\rho}^2}{(1.96\sqrt{\Psi_{ii}})^2} \quad (20)$$

If traffic is in a free flow state such that $S_e(\bar{\rho}) < R_e(\bar{\rho})$, we have:

$$\begin{aligned} \Delta t &\leq \frac{\bar{\rho}^2 \ell_i^2}{7.68 \bar{c}^2 v_f \bar{\rho}} \\ &= \frac{\bar{\rho} \ell_i^2}{7.68 \bar{c}^2 v_f} \geq \frac{1}{8} \frac{\bar{\rho} \ell_i}{\bar{c}^2} \frac{\ell_i}{v_f} \end{aligned} \quad (21)$$

The other issue of concern is avoiding densities within the 95% confidence interval that would be greater than jam density ρ_{jam} . This would require $\bar{\rho} + 1.96\sqrt{\Psi_{ii}\Delta t} \leq \rho_{\text{jam}}$, such that:

$$\Delta t \leq \frac{(\rho_{\text{jam}} - \bar{\rho})^2}{(1.96\sqrt{\Psi_{ii}})^2} \quad (22)$$

If traffic is in a congested state such that $S_e(\bar{\rho}) > R_e(\bar{\rho})$, we have:

$$\Delta t \leq \frac{(\rho_{\text{jam}} - \bar{\rho}) \ell_i^2}{7.68 \bar{c}^2 \omega} \quad (23)$$

In general, the recommendation in 23 is a more relaxed bound than in 21 as it is often

true that $v_f > \omega$. If one wanted to guarantee with 95% confidence that density values around the mean density of $\bar{\rho} = \rho_{crit}$, which is the upper bound of density values which result in $S_e(\bar{\rho}) > R_e(\bar{\rho})$, it would be required that:

$$\Delta t \leq \min\left(\frac{1}{8} \frac{q_{max}}{v_f} \frac{\ell_i^2}{\bar{c}^2 v_f}, \frac{\ell_i}{v_f}\right) \quad (24)$$

Of course, this cannot guarantee all traffic densities to be nonnegative.

Numerical Example

The purpose of this example is to show how we implemented the Gaussian approximation model in Algorithm 2 with a 95% confidence interval as allowed by the computation of the covariance matrix. To exhibit queue build-up and discharge by the model, we shall consider a large inlet cell arrival rate $A_{max}(t)$. In actual implementation of the model, the arrival flow rate varies with real-time data calibration. Consider an arterial road with a traffic signal that turns red, $g(t) = 0$ for $t \in [30, 150)$ and is green otherwise $g(t) = 1$ is modeled for 250 s. We choose to model this as a simple two cell setting with a cell length of .05 km. Assume a triangular fundamental relationship with free-flow speed $v_f = 100$ km/h, maximum flow rate $q_{max} = 2,100$ vph, backward wave propagation speed $\omega = 14$ km/h, and jam density $\rho_{jam} = 171$ veh/km, coefficient of variation $\bar{c} = 1$ for both states, maximum arrival rate $A_{max} = 2,100$, and a time step of $\Delta t = .045$ s. Then the flux function is calculated as:

$$\lambda(\bar{y}(.5, t)) = \min\{A_{max}(t), 14(171 - \bar{\rho}(1, t))\} \quad (25)$$

$$\lambda(\bar{y}(1.5, t)) = \min\{100\bar{\rho}(1, t), 2100, 14(171 - \bar{\rho}(2, t))\} \quad (26)$$

$$\lambda(\bar{y}(2.5, t)) = \min\{100\bar{\rho}(2, t), 2100g(t)\} \quad (27)$$

$$B = \begin{bmatrix} 20 & -20 & 0 \\ 0 & 20 & -20 \end{bmatrix} \quad (28)$$

$$\Gamma(t) = \begin{bmatrix} \sqrt{\lambda(\bar{y}(.5, t))} & 0 & 0 \\ 0 & \sqrt{\lambda(\bar{y}(1.5, t))} & 0 \\ 0 & 0 & \sqrt{\lambda(\bar{y}(2.5, t))} \end{bmatrix} \quad (29)$$

$$D = \frac{1}{20} \begin{bmatrix} \frac{\partial \lambda(\bar{y}(.5, t))}{\partial \bar{\rho}(1, t)} - \frac{\partial \lambda(\bar{y}(1.5, t))}{\partial \bar{\rho}(1, t)} & -\frac{\partial \lambda(\bar{y}(1.5, t))}{\partial \bar{\rho}(2, t)} \\ \frac{\partial \lambda(\bar{y}(1.5, t))}{\partial \bar{\rho}(1, t)} & \frac{\partial \lambda(\bar{y}(1.5, t))}{\partial \bar{\rho}(2, t)} - \frac{\partial \lambda(\bar{y}(2.5, t))}{\partial \bar{\rho}(2, t)} \end{bmatrix} \quad (30)$$

Case Study

An arterial road case study was carried out on eastbound Highway 14 in Brookings, South Dakota, between 22nd Avenue and 25th Avenue (Figure 3). Data were collected from 7:30 to 8:00 a.m. and the traffic largely consisted of passenger vehicles, with minimal amounts of lane changing. The road section is homogeneous, which allows us to use a homogeneous conservation equation.

To model traffic conditions on the road, we need to fit the parameters of the triangular fundamental diagram in Figure 1 to traffic conditions specific to the road. We desire to update the model with traffic flow and density data, and hence need to measure traffic conditions upstream. As a measure of performance, we also need to measure the traffic conditions downstream to test model estimates. The maximum queue size is measured downstream to compare with model estimates of density. Density values are used to estimate the queue size.

A video camera recorded the eastbound flow of traffic downstream from the traffic signal at 22nd Avenue at a point at which vehicles have reached free-flow speed. The intention is to record the flow and density rates of traffic flow upstream. The density upstream was measured using the space–mean speed along with the relation $q = \rho v$ such that

$$\bar{\rho} = \frac{1}{\Delta t} \frac{\sum_{i=1}^N \tau_i}{\Delta x} \quad (31)$$

where τ_i represents the trajectory of vehicle i . Vehicle trajectory times were measured over the length of the upstream segment and segmented by the offset cycle times. Offset in this scenario is the travel time of a vehicle from the detector location to the downstream traffic light. Vehicle time trajectories allow density to be calculated over the entire cycle time, just as an inductance loop detector would record vehicle occupancy over an entire cycle length.



FIGURE 3 Daft Logic mapping 2014 (12).

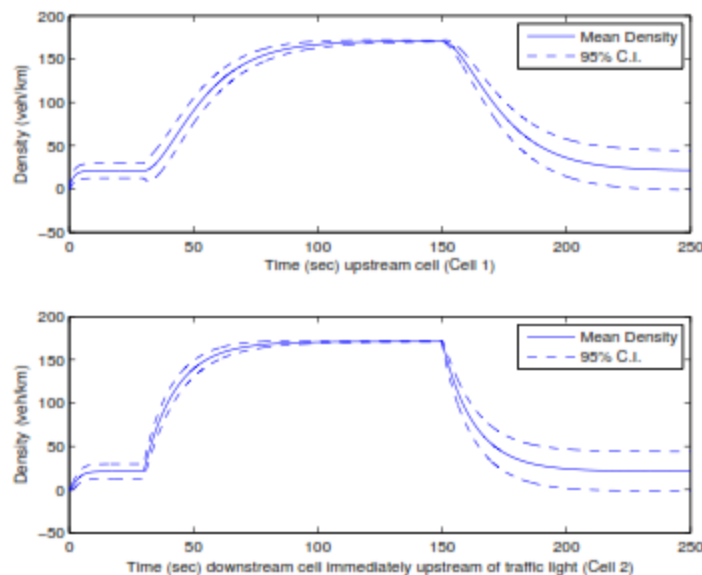


FIGURE 4 Traffic densities with 95% confidence intervals.

Flow rates are taken to be the vehicle count over the offset cycle time. A fixed point in the video is used to count vehicles with (Figure 4).

The Gaussian approximation model is a linear model, which allows it to be calibrated with real-time observations using the linear Kalman filter. This provides an advantage over other filters in that it is relatively fast. The filter is used in this case study to create optimal estimates of upstream density, $\rho(1,t)$, and inlet arrival flow rates, A_{\max} .

Another video camera recorded the downstream queue. This camera serves as the ground truth queue size to compare against the model's prediction. The model uses the upstream data to calibrate model estimates with the linear Kalman filter. The model then makes predictions of density for downstream cells. The density for each cell will then be used to determine if the traffic is considered to be in the queue. Cells are considered to be in the queue, starting from the most downstream cell, if the mean traffic speed is less than 8.05 km/h and the cell downstream is also in the queue.

Free-flow speed v_f , which is the slope of the free-flow regime in Figure 1, is assumed to be a bit larger than the speed limit: $56.32 + 3.21$ km/h. The capacity flow q_{cap} is determined from the saturation flow rate according to the procedure in Currin (13). The jam density ρ_{jam} is estimated by dividing the number of cars in a queue by the estimated length of the queue. The backward wave speed is derived as: $\omega = \frac{-q_{\text{cap}}}{\rho_{\text{jam}} - \rho_{\text{cap}}}$. This is the last of the parameters from the fundamental diagram that need to be determined.

The last model parameter we wish to find is the coefficient of variation of vehicle time headways, \bar{c} , from the fit of a lognormal distribution to time headways in both the congested and uncongested regimes. Data is fit to three-parameter lognormal distributions using statistical analysis system (SAS) with parameters in Table 2.

TABLE 2 Lognormal Fit to Time Headways

	Threshold	Scale	Shape	Mean	SD	\bar{c}	A^2 Test
Free flow	0.5869	0.6485	0.8867	3.420	3.097	0.906	0.4273
Congested	3.252	1.021	1.228	9.154	11.07	1.21	0.3707

NOTE: SD = standard deviation.

The null hypothesis is that the data are random samples from the assumed distribution. We use Anderson-Darling (A^2) goodness-of-fit test to test the alternative hypothesis, which is that the data are not from the assumed distribution. The A^2 test values in Table 2 are under their respective critical values for the 0.05-significance levels, thus the distributions fit well enough that we fail to reject the null hypothesis.

A time step of 1/2 of the recommendation of the Courant-Friedrichs-Lewy (CFL) condition is arbitrarily chosen as the largest time step which visually maintains stability (CFL condition is a necessary but not sufficient). A time step of this size does not guarantee nonnegative density estimates for the confidence interval. A smaller time step of 1/16 of the CLF recommendation, which maintains mostly nonnegative density confidence intervals, is shown for comparison.

The choice of cell size affects both accuracy and computation time. From an accuracy standpoint, smaller cell sizes increase the accuracy of the end of the queue. Cell sizes could be as small as the length of a vehicle. The total length of the observed section of road is approximately 248 m. The cell size was taken to be factors of the imperial road length.

In selecting the maximum queue from the model, the mean maximum queue size is used as the criteria. Cell densities consecutively above crawl density are considered for the maximum queue size confidence interval. Figure 5 represents an example of a case in which some cells would be considered part of the queue for the confidence interval, but not for the mean, and vice versa. The last five cells would be used in calculation of the maximum density, whereas only the last four cells would be considered for the lower confidence interval. Scenarios arise in which no density for the lower confidence interval are included in the queue size estimate, in which case the lower confidence interval for the maximum queue size is 0 vehicles.

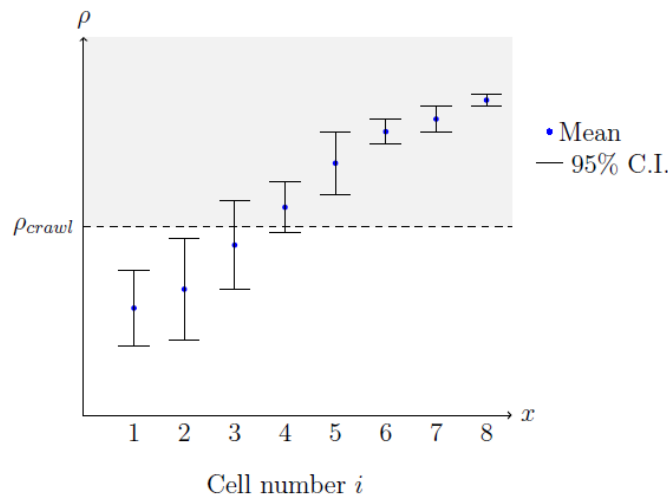


FIGURE 5 Densities in gray-shaded region considered for maximum queue size.

The queue size is determined to be the number of consecutive vehicles from the stop line traveling slower than crawl speed (taken to be $v_{\text{crawl}} = 8.05$ km/h). Since the model produces estimates of density, cells are considered to be in the queue if their density is greater than the crawl density. Crawl density is determined to be where the lines $q_{\text{crawl}} = v_{\text{crawl}}\rho_{\text{crawl}}$ and $q = \omega(\rho_{\text{jam}} - \rho)$ intersect.

Minimal lane changing occurs in the data set, thus the lanes are independent of each other and can be modeled individually. The queue sizes for the median lane were too small to determine model parameters from and thus are excluded in estimation. Queue sizes for the shoulder lane range from 0 to 10 vehicles.

RESULTS

Results of the model estimates are shown in [Figure 6](#) against the measured queue size. The model uses the first three iterations as a warm-up period to adapt to inlet flow and density values, and these are excluded from the figure. Since the time step in equation 20 requires knowledge of $\bar{\rho}$ and Ψ , it is difficult to choose a time step greater than zero that will guarantee nonnegative density estimates.

Additionally, we would not expect a confidence interval that is small enough to avoid nonnegative densities to also include measured conditions downstream. One purpose of developing a 95% confidence interval is to capture the range of values that could be expected in the downstream queue.

It is notable that the measurements are close to this range of values. It is also notable how much difference exists in model success for varying cell lengths and time step. Essentially confidence intervals can be constructed for a model with equal cell lengths which produce very different results, as is seen between [Figure 6a](#) and [b](#). This is because the stochastic model converges to the mean for smaller time steps, and the deviance of the two is what characterizes the 95% confidence interval. Choice of cell size affects the confidence interval as well. The time step decreases by the square of the cell length as the cell length decreases. This results in a conundrum in which a cell length is desired to be small enough to accurately capture the edge of the queue, but large enough such that the confidence interval provides usable information.

The measured observations are not all within the 95% confidence interval for any one model. We tested crawl speeds of 8.05 ± 3.22 km/h to determine if discrepancy existed for determining queue size during observation. Observed vehicles were only considered to be part of the queue if their speed was thought to be less than $v_{\text{crawl}} = 8.05$ km/h; however, this was not directly calculated. It could be possible that a vehicle was not observed to be in the queue, but would have been modeled. Results for the model estimate with crawl speeds of 8.05 ± 3.22 km/h did not improve the model's success.

Another possible explanation is error in data collection. While minimal lane changes occur in the videos, a short section of the road between the cameras' view was not observed. Unaccounted lane changes would affect queue size. In estimating the saturation flow rate, few queues were observed with the recommended initial queue of 10 vehicles. Thus the saturation flow rate could be inaccurate. Additionally, this would impact the value of the backwards wave speed ω . The degree of accuracy in measurements of distance and time in videos can affect the accuracy of the all of the data as well.

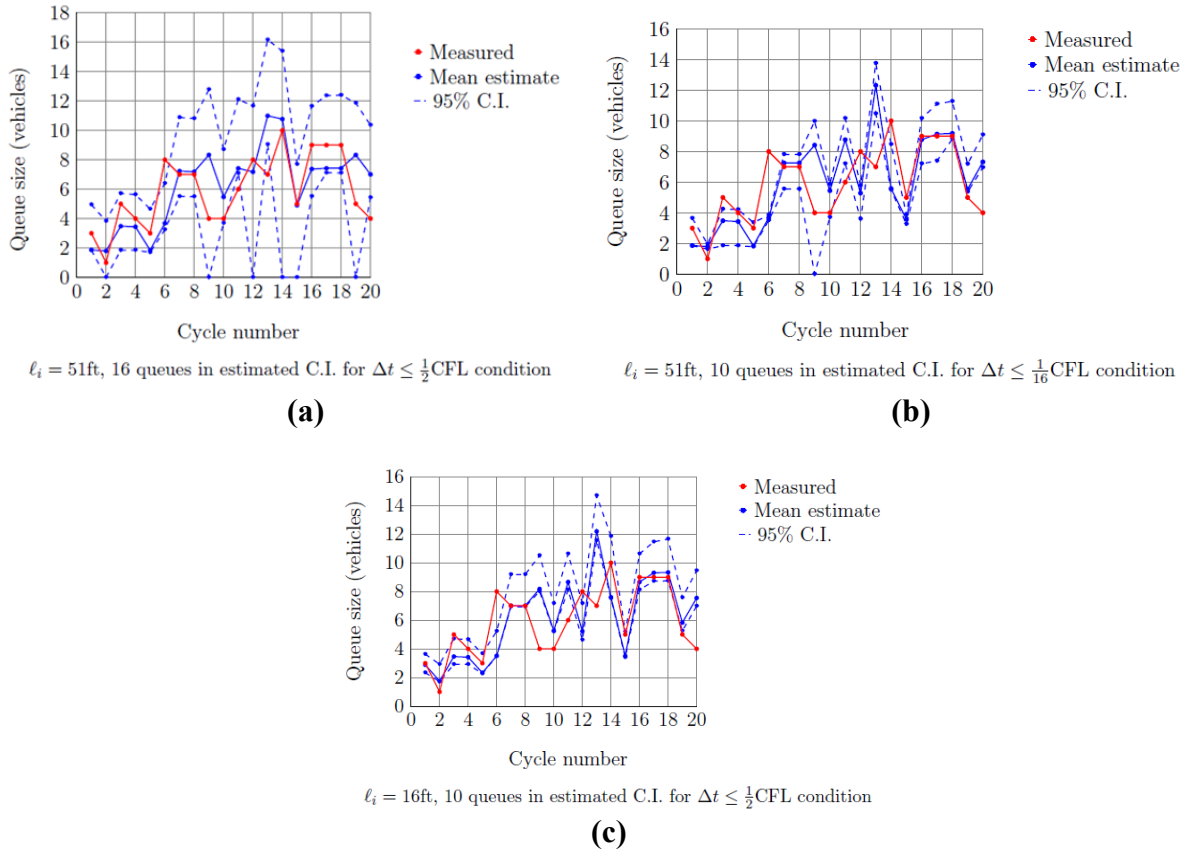


FIGURE 6 Comparison between estimated and measured queue sizes.

Throughout this case study, many updates were made to the data to improve their accuracy. The changes that the model was most sensitive to were changing the upstream flow rates and changing the free-flow speed and backward wave speed. This emphasizes the importance of fitting the fundamental parameters to relevant historical data. Updating density values resulted in negligible change, as is expected for a Markov model. Updating flow values greatly affected model predictions. This shows the importance of updating the inlet flow rate, as the Kalman filter assigns the maximum flow allowed into the road to be the inlet flow rate until the next filter iteration. Updates to density produce temporary effects in the model estimates, and was not as important in updating the model.

CONCLUSION

Macroscopic models accumulate random microscopic phenomenon across lengths of roads. Stochastic models have been developed to simulate the stochastic nature of vehicle interactions and arrivals. In order to produce computationally viable models, researchers have added Gaussian noise to a deterministic model to create a stochastic model. This could lead to negative estimated densities, which is an undesirable physical feature. The model outlined in Jabari and Liu (6) creates a doubly stochastic Poisson counting process based on stochastic time headways. This model was developed to create a stochastic cell transmission model with the limiting

features of the stochastic process to be the deterministic Godunov scheme. The deviance between the stochastic and deterministic model is derived as a stochastic differential equation, from which the variance of the mean counting process is derived. This stochastic differential equation is shown to be a Gaussian distribution about the fluid limit.

This paper implements the Gaussian approximation model proposed in Jabari and Liu (7) on an arterial road in Brookings, South Dakota. The deterministic model is computationally viable, its variance is controllable such that nonnegative density distributions are obtainable, and the mean of the original stochastic model used in the derivation of the Gaussian approximation converges in the fluid limit to Godunov scheme dynamics. The linearity of the model allows for the use of Kalman filtering for calibrating the model to real-time data.

The case study was able to produce results with 80% accuracy in predicting observed queue sizes within the confidence interval at the loss of guaranteeing nonnegative densities. In practice, it does not seem reasonable to expect that a model maintaining nonnegative densities with 95% confidence produces a large variance. It must be determined which takes priority in modeling traffic flow: providing ranges of traffic flow values that accurately represent the modeled situation, or maintaining physical phenomenon that densities should be nonnegative.

FUTURE WORK

The Gaussian approximation model can be applied to any traffic operations management systems that require traffic predictions. This includes advanced signal timing, ramp metering, and advanced wait time notice.

The CTM was developed to solve the LWR model for complicated network topologies. Thus the Gaussian approximation could be investigated and developed for these scenarios. To do this, one would derive this from deterministic merge and diverge dynamics.

It would be of interest for those implementing this model in the future to use Monte Carlo simulation of the stochastic process to check whether the 95% confidence interval about the mean is valid for various time steps. The authors of this paper assumed the derivation of the variance of the stochastic process to its mean to be sufficient.

REFERENCES

1. Cambridge Systematics, Inc., and Texas Transportation Institute. *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation*. Technical report. FHWA, U.S. Department of Transportation, 2005.
2. Lighthill, M. L., and G. Whitham. On Kinematic Waves. II: A Theory of Traffic Flow on Long Crowded Roads. *Proc., Royal Society of London: Series A: Mathematical and Physical Sciences*, 1955, pp. 317–345.
3. Richards, P. I. Shockwaves on the Highway. *Operations Research*, Vol. 4, 1956, pp. 42–51.
4. Kim, T., and H. M. Zhang. A Stochastic Wave Propagation Model. *Transportation Research Part B*, Vol. 42, No. 7, 2008, pp. 619–634.
5. Sumalee, A., R. X. Zhong, T. L. Pan, and W. Y. Szeto. Stochastic Cell Transmission Model: A Stochastic Dynamic Traffic Model for Traffic State Surveillance and Assignment. *Transportation Research Part B*, Vol. 45, No. 3, 2011, pp. 507–533.
6. Jabari, S., and H. Liu. A Stochastic Model of Traffic Flow: Theoretical Foundations. *Transportation Research Part B*, Vol. 46, No. 1, 2012, pp. 156–174.

7. Jabari, S., and H. Liu. A Stochastic Model of Traffic Flow: Gaussian Approximation and Estimation. *Transportation Research Part B*, Vol. 47, 2013, pp. 15–41.
8. Daganzo, C. The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research Part B*, Vol. 28, No. 4, 1994, pp. 269–287.
9. Daganzo, C. The Cell Transmission Model: Network Traffic. *Transportation Research Part B*, Vol. 29, No. 2, 1995, pp. 79–93.
10. Godunov, S. A Difference Scheme for Numerical Solution of Discontinuous Solution of Hydrodynamic Equations. *Matematicheskii Sbornik*, Vol. 89, No. 3, 1959, pp. 271–306.
11. Leveque, R. J. *Numerical Methods for Conservation Laws*, Vol. 132, Springer, 1992.
12. Daft Logic. Distance Calculator. 2014. Available at <http://tinyurl.com/pdudavo>. Accessed June 2014.
13. Currin, T. R. *Introduction to Traffic Engineering: A Manual for Data Collection and Analysis*. Brooks/Cole, 2001.

POSTER PRESENTATION

**Investigation of Performance and Lane Utilization
Within a Passing Lane on a Two-Lane Rural Highway**

ZACHARY FREEDMAN
Sabra, Wang & Associates

AHMED AL-KAISY
Montana State University

An investigation into platooning and passing maneuvers within a passing-lane section on a rural two-lane, two-way highway is presented in this study. The study site is located on US-287 between the town of Townsend and the city of Helena in the state of Montana. Traffic volumes at study site, while considered relatively low, are typical on two-lane highways in many rural states. Per-lane analysis of performance measures and lane utilization (volume split) were used to indirectly examine passing maneuvers and lane changing at successive locations within the passing lane section. For the case study site, it was evident that traffic performance became relatively stable beyond 0.5 mi into the passing lane for the traffic volumes investigated. Therefore, results strongly suggest that most passing maneuvers already took place before the 0.5-mi station and that the actual passing lane length is well beyond the optimal length required for breaking up platoons and improving performance.

INTRODUCTION

Two-lane highways constitute the vast majority of highway facilities particularly in states where most vehicular travel takes place in rural areas. Passing maneuvers are restricted on those highways and are typically performed using the opposing lane when sight distance and gaps in the opposing traffic stream permit. This has serious implications on traffic performance and safety. From traffic performance perspective, the limited passing opportunities would result in higher impact of slow-moving vehicles on mobility and performance. This impact generally increases with the increase in traffic level in the two directions of travel, the proportion of slow-moving vehicles in the traffic stream, and the average speed differential between the mainstream and slower vehicles. The use of passing lanes is known to alleviate this unique operational characteristic on well-traveled two-lane highways. Passing lanes allow vehicles traveling at faster speeds to overtake slow-moving vehicles, thus breaking up platoons and reducing delays associated with inadequate passing opportunities. Therefore, it is well known that using passing lanes generally improves operations and results in better level of service on two-lane highways. Two very important aspects in planning and designing passing lanes on two-lane highways are the determination of the appropriate passing lane length and the frequency at which passing lanes are provided. This study examines platooning and performance within the passing lane, which is important in understanding the length requirement for passing lanes on two-lane highways.

BACKGROUND

In the course of the literature review performed in this study, only limited research was identified that is related to passing-lane length requirement. In this section, information available about passing lane length requirement in the current practice as well as the few relevant studies in the literature are presented.

The *Highway Capacity Manual* (HCM) (1) recommends a passing-lane length based on traffic level as shown in Table 1. The source of the values provided in this table is a study that was conducted more than 25 years ago using traffic simulation (2).

The *Policy on Geometric Design of Highways and Streets* (3), known as the Green Book, suggests values for passing lane length that are very consistent with the values in the current HCM. The absolute minimum length for passing lanes according to the Green Book is 300 m (1,000 ft) excluding tapers. Also, the optimal length provided in the green book is 0.8 to 3.2 km (0.5 to 2.0 mi) which is the same range of values provided in Table 1.

A few researchers have investigated performance or passing maneuvers within passing lanes with the objective of understanding the passing lane length requirement under various traffic conditions.

Harwood and Hoban (4) used the traffic simulation model TWOPAS in an attempt to determine optimal length and spacing of passing lanes. Their data showed that optimal passing lane length was dependent on traffic volume. For low traffic level, the optimal passing lane length is 0.5 to 0.75 mi. For medium traffic level (around 400 vph) the optimal length is .75 to 1.0 mi, and for higher traffic level (equal or greater than 700 vph) the optimal length is 1.0 to 2.0 mi. Results from this study were used in developing the HCM recommended values for passing lane lengths in the HCM 2000 and HCM 2010 (1, 5).

Another study by Harwood and St. John (2) investigated 12 passing lane sites in six states. The purpose of this study was to determine the effective length and spacing of passing lanes. At each site, automatic traffic recorders and observers were used to collect data. The data recorders collected vehicle mix, speed, headway, and acceleration while the observers primarily counted the passing maneuvers. The locations of the data recorders were 500 ft upstream of the entrance taper, 100 ft downstream of the entrance taper, in the middle of the passing lane, 300 ft from the beginning of the exit taper, 500 ft downstream of the exit taper, and 1 mi downstream of the exit taper. Six hours of data were collected at each site. The main measures of effectiveness for this study were traffic speed, percent of vehicles in a platoon, and passing rate. The study found that operational benefits of passing lanes exist over a distance of three to 8 mi and that the optimal length of passing lanes generally vary from 0.5 to 1.0 mi.

TABLE 1 Optimal Passing Lane Length in the HCM

Directional Demand Flow Rate (pc/h)	Optimal Passing Lane length (mi)
≤ 100	≤ 0.5
$> 100 \leq 400$	$> 0.5 \leq 0.75$
$> 400 \leq 700$	$> 0.75 \leq 1.0$
≥ 700	$> 1.0 \leq 2.0$

NOTE: pc/h = passenger cars per hour.

Woolridge et al. (6) researched the optimal length and spacing of passing lanes on two-lane highways using the traffic simulation model TWOPAS. A generic two-lane highway with varying length of passing lane was tested. The spacing between passing lane sections was varied between 1 and 8 mi at 1-mi intervals and the length of passing lane was varied between .25 and 2 mi at .25-mi intervals. The results of the simulation showed that the optimal length of passing lane varied between 0.8 and 2.0 mi and the optimal spacing varied between 3.5 and 11 mi.

Gattis et al. (7) studied sites with a 2+1 passing-lane configuration (also known as a three-lane configuration or an alternating passing lane) in Arkansas. This configuration has one lane in each direction (like a normal two-lane highway) and center lane that acts as a passing lane alternating in direction. This is demonstrated in [Figure 1](#).

Four sites were selected for study of platoon-breakup and passing rates: two on US-70, one on US-65, and one on US-82. Study results showed that the greatest reduction in platooning was between 0.93 and 1.86 mi into the three-lane section and that the platooning at the beginning of the three-lane segment was always greater than at the end. Further, the researchers found that the maximum passing rate occurred between 0.93 and 1.86 mi into the site. Past 1.86 mi, passing rate leveled off or declined. Passing rates also increased with higher volumes.

May and Emoto (8) conducted a field study to analyze the passing lane usage. A plot of the total volume versus the passing lane volume was created. A best-fit line was developed and a passing experiment was conducted to determine the number of passes vehicles made per mile of passing lane. The number of passes varied from 1.5 passes per mile to 1.75 passes per mile.

PROBLEM STATEMENT

The periodic use of passing lanes is one of the most common operational improvements on two-lane highways. In many instances, it is proven more feasible than other costlier improvements such as the super 2+1 and the upgrade to a four-lane highway facility. In planning and designing passing lanes, two parameters are deemed very important; the passing lane length required and the interval at which the passing lane should be provided on a particular highway to achieve the desired performance improvement. This current research is associated with the former of the two design parameters.

The determination of the needed or optimal passing lane length requires a good understanding of traffic behavior within the passing lane sections. Developing such an understanding is somewhat challenging in that it should largely rely on empirical field observations, particularly in the absence of advanced and proven simulation models. Vehicular

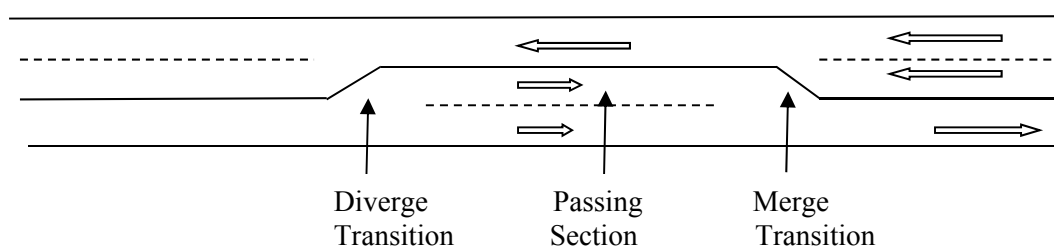


FIGURE 1 “2+1 Passing Lane” configuration.

interactions in most simulation models are based on highly theoretical and sometimes very simplistic car-following and lane-changing models that may not yield realistic results. This is especially true for the simulation of two-lane highway operations where only a limited number of simulation models are available to researchers and practitioners. On the other hand, field investigations are constantly challenged, to various degrees, by limitations to site geometrics, traffic conditions, data collection techniques, or combinations of the aforementioned factors. Nonetheless, field observations offer valuable information about the dynamics of vehicular interactions within the passing lane section that is much needed in the literature. As such, it is the objective of this research is to empirically explore the dynamics of platoon break up within a typical passing-lane section with respect to the passing-lane length requirement. While this case study involved a single passing lane site, findings of the case study is deemed applicable to many similar sites on two-lane highways in rural areas.

STUDY SITE

The study site investigated in this research is located on US-287 Southbound between the town of Townsend and the city of Helena in the state of Montana. This particular passing lane is roughly located at around 8 mi north of Townsend and 25 mi south of Helena. The passing lane is around 2.33 mi in length and is located in level terrain. The highway has standard lane width and wide shoulders at the passing lane location. No major access points exist which could affect travel patterns within the passing lane of interest. Only one unpaved driveway in the northbound direction exists and is located further downstream in the study area. Southbound traffic using the driveway was observed in the field and found too small to have any noticeable influence on lane utilization patterns. The annual average daily traffic at this site in the data collection year (2008) was around 5,000 vehicles per day.

STUDY DESIGN

In order to investigate performance and lane utilization, multiple detector stations were used at the study site. Specifically, five automatic traffic recorder stations were used to gather data both upstream and within the passing lane over a 1-week period in June 2008. As illustrated in [Figure 2](#), the first station was located 250 ft upstream of the passing lane; the other four stations were located at 0.25, 0.5, 1.0, and 1.5 mi downstream of the lane addition taper into the passing lane. Traffic data including counts, speeds, headways, and vehicle classification were obtained at the single lane upstream of the passing lane and for each individual lane within the passing lane section. Counts and vehicle classification were used to establish lane utilization within the passing lane while headways, speeds, classification, and counts were used in investigating performance upstream and within the passing-lane section. Three different performance measures were used in the study; percent followers, average travel speed, and follower density. Percent followers represents the percentage of vehicles with short headways (less than 3 s) in the traffic stream. This is the surrogate measure used by the HCM for estimating percent-time-spent following in the field. Follower density is the number of followers in a directional traffic stream over a unit length, typically 1 km or 1 mi. It can easily be found by multiplying percent followers and traffic density. This measure, which has been used in South Africa (9), was found the most

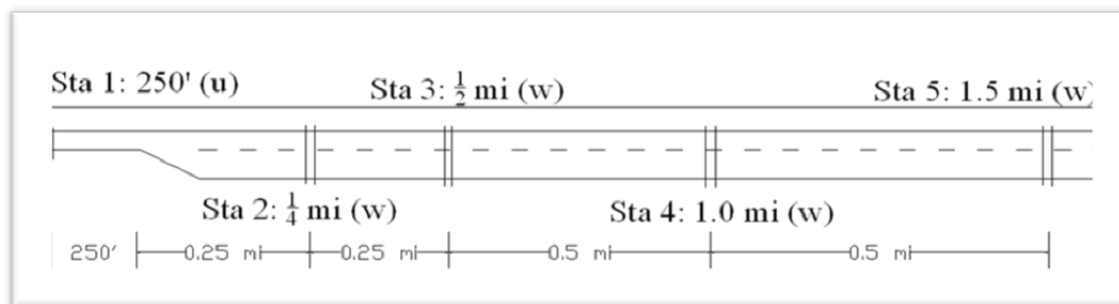


FIGURE 2 Field data collection setup.

promising among other measures investigated in a recent study (10). Average travel speed is the arithmetic mean of all speeds observed at a point location and is the secondary performance measure used by the HCM for Class I two-lane highways, which includes rural two-lane arterials and primary highways (1).

STUDY RESULTS

Per Lane Examination of Performance Measures

In order to understand traffic behavior within the passing lane, performance measures most relevant to platooning and impedance were examined individually on the normal and passing lanes. As discussed earlier, three measures of performance were examined in this analysis; percent followers, follower density, and average travel speed. In this analysis, plots of each performance measure were created as a function of distance into the passing lane at different traffic levels. Figure 3 shows the percent followers at successive data collection stations within the passing lane section at three different traffic levels.

On the normal (left) lane, the percent followers seems to rise significantly in the first 0.25 mi, then drops to or slightly below its original level (upstream of passing lane) between 0.25 and 0.5 mi on that same lane and finally levels off. On the passing (right) lane, the changes exhibit a different pattern: percent followers remains relatively steady in the first 0.25 mi, slightly drops between 0.25 and 0.5 mi, and eventually levels off. These patterns are common to all traffic levels investigated. One observation that is somewhat unexpected here is the relatively low reduction in percent followers in the passing lane section despite the fact that the total traffic on the single lane has split over two lanes within the passing section. In fact, percent followers at 0.25 mi station is much higher than the original value when averaged over the two lanes. It should be understood that those results do not suggest deterioration in performance upon entering the passing lane, but rather a sign of the limitation in using the time headway as the only performance indicator.

Figure 4 shows the follower density at successive data collection stations within the passing lane section at the same different traffic levels. As shown, follower density in each lane drops from the pre-passing-lane value within the first 0.25 mi, in spite of the rise in percent followers. This is because traffic volume splits into two separate volumes in the two available

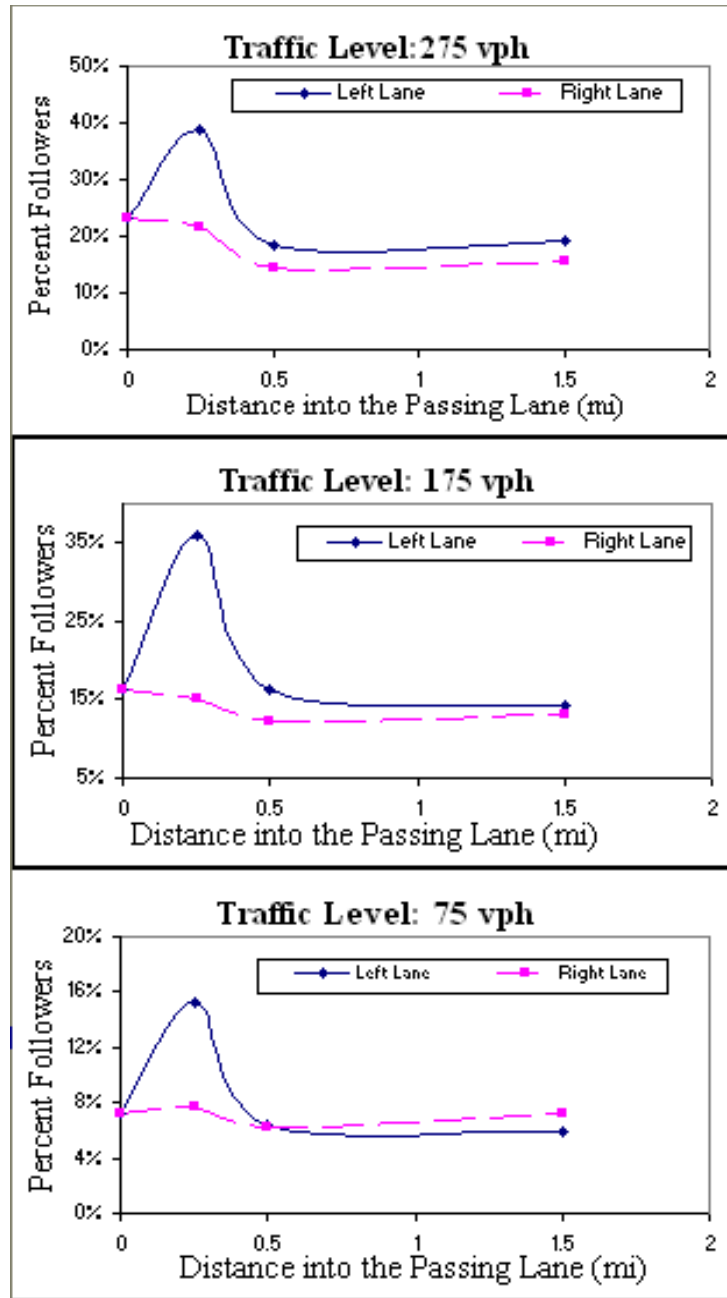


FIGURE 3 Percent followers at the passing lane section for various traffic levels.

lanes beyond the taper. Apparently, the effect of the decrease in volume on follower density is higher than that of the increase in percent followers at 0.25 mi station, which was clearly exhibited in Figure 3. At 0.25 mi into the passing lane, follower density in the left lane is significantly higher than that in the right lane. This may well be related to the notable increase in percent followers at this station as illustrated in Figure 3. The only exception is when traffic level is at around 75 vph. At this traffic level, it seems that the density in the normal lane has decreased to an extent that caused follower density on the left lane to be lower than that on the

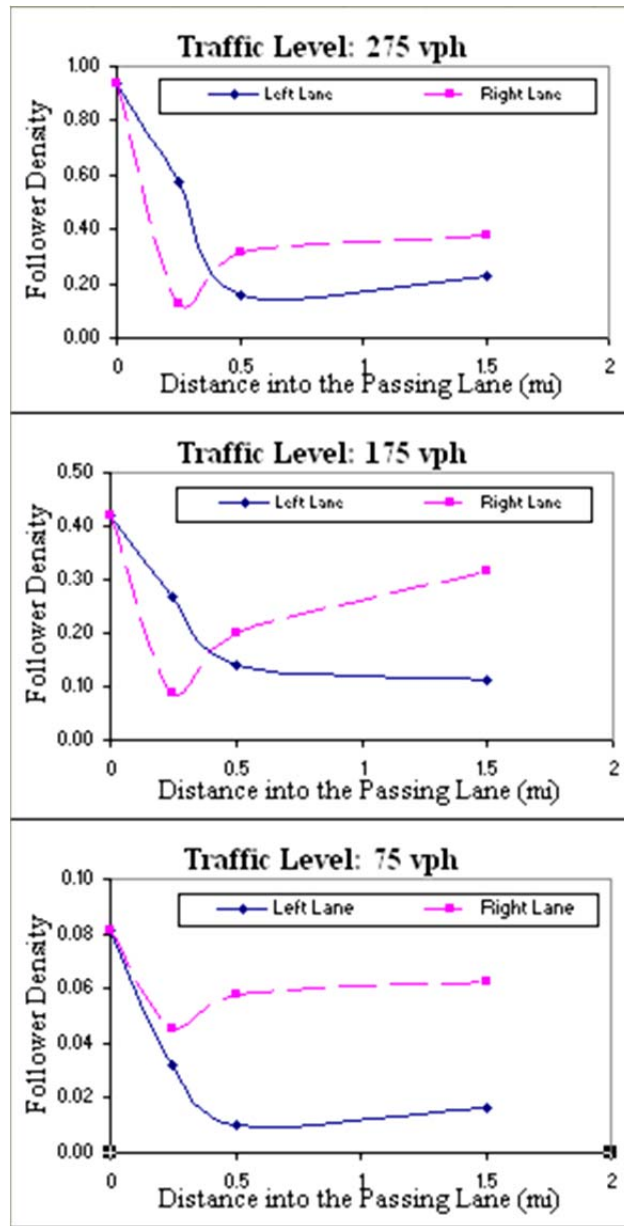


FIGURE 4 Follower density by lane at the passing lane section.

right lane regardless of the increase in percent followers. At 0.5 and 1.5 mi into the passing lane section follower density is always higher on the passing lane than on the normal lane. In general, the lower the traffic level, the higher the discrepancy in follower density between the two lanes. Given the relatively close values for percent followers at the 0.5- and 1.5-mi stations, results shown in this figure indicate that more traffic use the right lane in general, including free-moving vehicles outside vehicular platoons which constitute high proportion of vehicles at low traffic levels.

Figure 5 shows average travel speed on the two lanes within the passing section at various traffic levels. Speed changes along the distance into the passing lane exhibit very

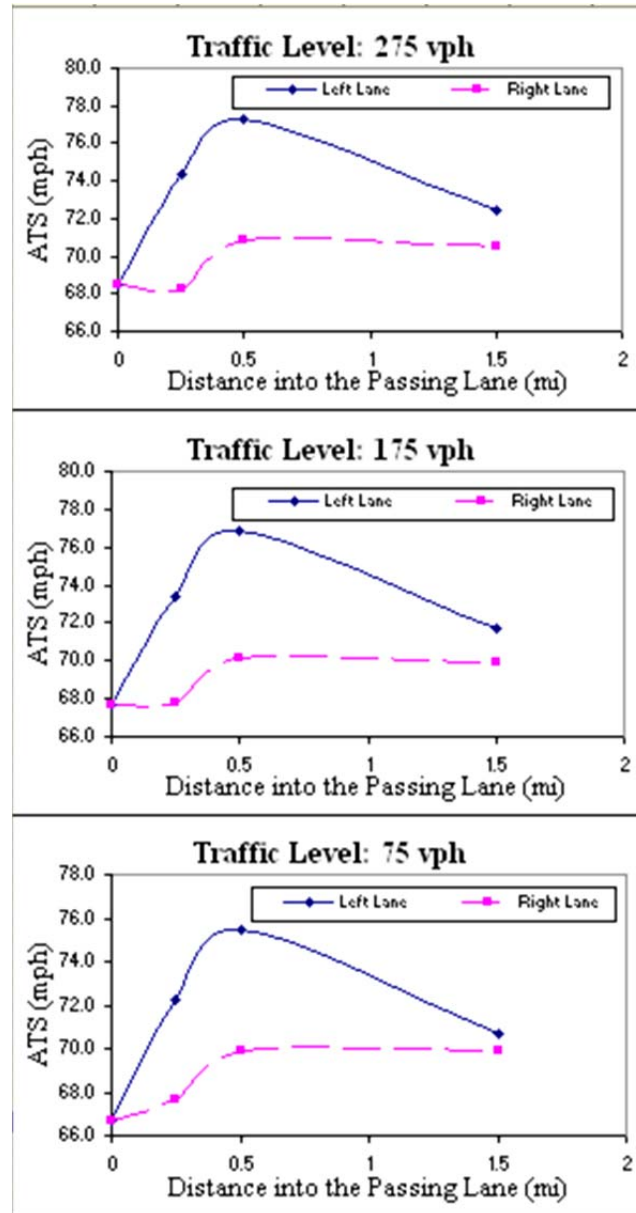


FIGURE 5 Average travel speed at the passing lane section.

consistent trends at all traffic levels particularly in the normal (left) lane. In this lane, speed increases sharply in the first 0.5 mi (around 13% increase), then declines to a lower value at 1.5 mi into the passing lane. The final speed at 1.5-mi distance is still notably higher than the original pre-passing lane speed. Specifically, the speed at 1.5-mi distance is roughly the middle value between the pre-passing lane speed and that at 0.5-mi distance into the passing lane. A totally different trend is shown for the other lane. In the passing lane, speed generally maintains its pre-passing lane value with the exception of a small increase (around 2 mph) between 0.25 and 0.5 mi into the passing lane. As expected, the final speed in the passing lane at 1.5 mi distance is lower than its counterpart for the normal lane, but is still higher than the pre-passing lane speed.

The previous results reveal some of the important aspects of drivers' behavior in using passing lanes. Upon arriving at the passing lane section, traffic is normally split into two streams, one in each travel lane. Examining performance measures in Figures 3 to 5 suggest that aggressive drivers always opt to remain in the normal (left) lane, while slower-moving vehicles and other vehicles that are not much affected by slower speeds tend to use the passing lane. The sharp increase in short headways in the normal lane (as evidenced by the percent followers at 0.25 mi in Figure 3) and the notable increase in speed (Figure 5) both suggest that the majority of aggressive drivers use the left lane upon arriving at the passing-lane section. On the other hand, changes on the passing lane are less notable. Percent follower slightly declines and average speed slightly increases (around 4%) mainly because of the drop in traffic flow from the pre-passing lane level.

Per Lane Analysis of Traffic Volumes

To gain further insights into traffic behavior within the passing lane section, and to better understand the performance trends discussed in the previous section, traffic counts at the successive data collection stations are analyzed at various traffic levels and results are presented in Figure 6. This figure shows traffic volume split in the two available lanes as a percentage of the total volume at the data collection stations. As shown in this figure, the majority of traffic chose to use the passing lane within the distance investigated. The analysis shows that, at 0.25-mi distance, between 60% and 80% of traffic utilizes the right (passing) lane depending on the traffic level. At 0.5-mi distance into the passing lane, these percentages increase and they all fall roughly in the range of 70% to 85%.

This indicates that many drivers may choose to change lanes beyond the taper section and well into the passing-lane section. One possible explanation could be as follows: some moderate (nonaggressive) drivers stuck behind slow-moving vehicles may choose to take the right lane, but only after passing the vehicle(s) they were impeded by, while others may make the same decision when caught up by aggressive drivers traveling at higher speeds in the left lane. Besides the increase in passing lane split, this hypothesis is also supported by the increase in average travel speed and the slight decrease in percent followers in the passing lane due to these moderate drivers moving into the passing lane.

At 1.5-mi distance, the passing-lane split slightly declines and varies in the range between 65% and 75% depending on the traffic level. This indicates that some drivers may decide to return to the normal lane well in advance of the passing-lane end. This is particularly true for drivers who use the right lane mainly to allow passing of faster vehicles and not because they drive at very low speeds. The other major trend exhibited in this figure is the increase in the passing lane traffic split at lower traffic levels and vice versa, which is somewhat expected.

Interpretation of Results in Regard to Driver Behavior

In this section, the trends exhibited in the previous figures are analyzed in relation to drivers' lane selection (lane changing) behavior within the passing-lane section. Those figures show that the percentage of short headways in the normal lane rises sharply in the first 0.25 mi. Half mile into the passing section, the percentage of short headways in the normal lane drops to below the pre-passing lane value. Meanwhile, average travel speed of vehicles in this same lane continues

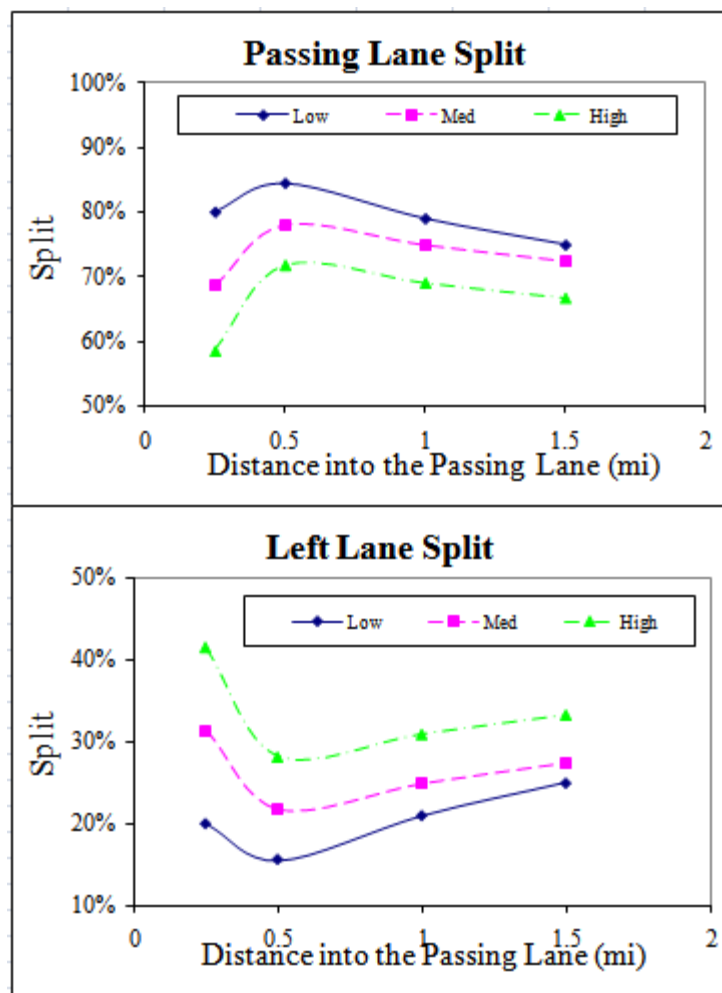


FIGURE 6 Volume split over available lanes at the passing-lane section for various traffic levels.

to increase (until the 0.5-mi station). Those trends can be interpreted as follows: between the upstream and the 0.25-mi stations aggressive drivers have chosen the left lane and begun speeding up. The moderate drivers (or majority of them) also chose the left lane in order to pass the slowest vehicles before they move to the passing lane. Between 0.25 and 0.5 mi, the moderate drivers shift to the right lane upon passing the slower vehicles. This is supported by the dramatic drop in percent followers in the left lane. Further, Figure 5 shows the highest percent of drivers in the passing lane occurring at the 0.5-mi station. By the time the traffic reaches the 0.5-mi station, the aggressive drivers have been able to pass the moderate (and slowest vehicles) and the moderate drivers have been able to pass the slowest vehicles. Beyond that point, drivers begin transitioning back gradually to the normal lane from the passing lane. This implies that, for this particular site, a passing lane length between 0.75 and 1.0 mi may prove to be adequate for all passing purposes.

Implication of Results on Passing Lane Length

In this research, passing maneuvers and lane changing were indirectly examined through the use of performance measures and volume lane split at successive locations within a passing lane section. Traffic volumes at the study site, though ranged between low and moderate traffic levels, are typical on two-lane highways in many rural states. For this case study site, it was evident that traffic performance became relatively stable beyond half a mile into the passing lane, which suggests that most passing maneuvers already took place before the 0.5-mi station. Accordingly, it is logical to conclude that the actual length of the passing lane under study is well beyond the optimal length given traffic conditions prevalent at the study site.

CONCLUSION

This paper presents a case study investigation of traffic performance and lane utilization at a passing-lane section on a rural two-lane highway in Montana. The analyses provided valuable insights that led to better understanding of driver's lane selection and lane changing behavior within the passing lane section. While this case study involved a single passing lane site, the results presented in this paper are applicable to many similar passing lanes on rural two-lane highways.

REFERENCES

1. *Highway Capacity Manual*. Fifth Edition. Transportation Research Board of the National Academies, Washington, D.C., 2010.
2. Harwood, D. W., and A. D. St. John. Passing Lanes and Other Operational Improvements on Two-Lane Highways. Report No. FHWA-RD-85-028. FHWA, U.S. Department of Transportation, 1984.
3. American Association of State Highway and Transportation Officials. *A Policy on Geometric Design of Highways and Streets*. 6th Edition, AASHTO, Washington, D.C., 2011.
4. Harwood, D. W., and C. J. Hoban. Low-Cost Methods of Operational Improvements on Two-Lane Highways. Report No. FHWA/IP-87/2. Midwest Research Institute and FHWA, 1987.
5. *Highway Capacity Manual*. Fourth Edition. Transportation Research Board of the National Academies, Washington, D.C., 2000.
6. Woolridge, M., C. Messer and B. Heard. Design Criteria for Improved Two-Lane Section (Super 2). Texas Transportation Institute, Project 0-4064, 2002.
7. Gattis, J., R. Bhawe, and L. Duncan. Alternating Passing Lane Length. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1961, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 16–23.
8. Emoto, T. C., and A. D. May. Operational Evaluation of Passing Lanes in Level Terrain. Final Report. Report UCBRR-88-13. Institute of Transportation Studies, University of California at Berkeley, 1988.
9. Van As, C. The Development of an Analysis Method for the Determination of Level of Service of Two-Lane Undivided Highways in South Africa. Project Summary. South African National Roads Agency, Limited, 2003.
10. Al-Kaisy, A., and S. Karjala. Indicators of Performance on Two-Lane Rural Highways: An Empirical Investigation. In *Transportation Research Record 2071*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 87–97.

POSTER PRESENTATION

Urban Road Network Macroscopic Fundamental Diagram Analysis Under Vehicular Ad-Hoc Networks Environment

ZHE XU

BIN RAN

University of Wisconsin–Madison

PETER J. JIN

Rutgers University

This research aims to analyze the urban road network macroscopic fundamental diagram under vehicular ad-hoc networks (VANETs) environment. The macroscopic (network) fundamental diagram (MFD or NFD), which relates network space–mean density and flow, has been shown to exist in urban network with homogeneous traffic conditions. Although MFD represents the network performance and is an important indicator for traffic management, the practical application of the MFD is limited to the fact that the homogeneous traffic condition is generally not observed in the real network. However, with wireless communication technology, VANETs provide such environment and therefore could be evaluated by MFD. VANETs are networks in which each node is a vehicle. The adjacent vehicles in VANETs are equipped with wireless communication devices. Under VANETs environment, the road network system provides communications not only between individual vehicles, but also between vehicles and nearby fixed equipment and roadside facilities. Therefore, the vehicles in VANETs environment receive real-time road information, thus improve traveling safety and efficiency.

The primary object of this study is to explore whether a well-defined MFD exists in urban road network under VANETs environment. To analyze the MFD under VANETs environment, a simulation test is conducted with grid road network in NS-3 simulation platform. The MFD of the simulated network is presented in the study. The simulation results are compared between different wireless communication scenarios. The results show that the urban road network performs a MFD under VANETs environment. Meanwhile, wireless communications under VANETs environment has an impact on the shape of the MFD for the simulated road network.

INTRODUCTION

Ongoing Work

The MFD or NFD, which relates network space–mean density and flow, has been shown to exist in urban network with homogeneous traffic conditions. Although MFD represents the network performance and is an important indicator for traffic management, the practical application of the MFD is limited to the fact that the homogeneous traffic condition is generally not observed in the real network.

However, with wireless communication technology, VANETs provide such environment and therefore could be evaluated by MFD.

VANETs are networks in which each node is a vehicle. The adjacent vehicles in VANETs are equipped with wireless communication devices. Under VANETs environment, the road network system provides communications not only between individual vehicles, but also between vehicles and nearby fixed equipment and roadside facilities. Therefore, the vehicles in VANETs environment receive real-time road information, thus improving traveling safety and efficiency. Due to the advantages of VANETs, it has received much attention from the automotive industry, government agencies, and researchers.

The primary object of this study is to explore whether a well-defined MFD exists in urban road network under VANETs environment. To analysis the MFD under VANETs environment, a simulation test is conducted with gird road network in NS-3 simulation platform. The MFD of the simulated network is presented in the study. The simulation results are compared between different wireless communication scenarios.

LITERATURE REVIEW

VANETs Network and Vehicle Wireless Communications

Introduction to VANETs Network

VANETs are networks in which each node is a vehicle. The adjacent vehicles in VANETs environment are equipped with wireless communication devices and therefore they are no longer isolated systems. Such systems aim to provide communications between (1) individual vehicles and (2) between vehicles and nearby fixed equipment, or roadside units. The goal of VANETs is to improve traffic safety and by traffic efficiency by providing timely information to drivers and concerned authorities.

Wireless Communication Standards and Data in VANETs

The automotive industry is working to develop the dedicated short-range communication (DSRC) technology, for use in vehicle-to-vehicle and vehicle-to-roadside communication in VANETs. There are several standards which defines the wireless communication standards in VANETs environment. The IEEE 802.11p amendment for wireless access in vehicular environments (WAVE) defines the wireless communication band of DSRC in VANETs. The IEEE 1609.2, 1609.3, and 1609.4 standards define the security, network services, and multichannel operation standard in VANETs. The SAE J2735 Message Set Dictionary supports interoperability among DSRC applications through the use of standardized message sets, data frames and data elements.

Macroscopic Fundamental Diagram Analysis Under VANETs Environment: Ongoing Work

Simulation Study in NS-3 Simulation Platform

Proposed simulation test is conducted with the NS-3 network simulator platform. Vehicle

mobility and network communication are integrated through events under NS-3 simulator. User-created event handlers can send network messages or alter vehicle mobility each time a network message is received and every time vehicle mobility updated by the model.

Test Network

Proposed simulation network is a grid network which consists of grids and links with same properties. This study employs NS-3 network simulator as a simulation platform.

Simulation Setting

Simulation will be conducted with the NS-3 network simulator. Different simulation scenarios will be tested during simulation: (1) grid network simulation with communications between vehicles to vehicles (V2V) and vehicles to infrastructure (V2I) and (2) grid network simulation with only communications between V2V.

Macroscopic Fundamental Diagram Analysis in VANETs: Ongoing Work

MFD in Simulation Network

The MFD of the simulation network, which relates network flow and space-mean density, will be presented under VANETs environment according to the simulation results. The MFDs for different simulation scenarios will be compared and analyzed.

Impacts of Wireless Communication in VANETs to MFD

By comparing the MFDs for different simulation scenarios, the impacts of wireless communication in VANETs to MFD will be discussed.

CONCLUSION AND FUTURE WORK

The research investigates the urban road network macroscopic fundamental diagram under vehicular ad-hoc networks (VANETs) environment. The major contribution of this study is to show the existence of macroscopic fundamental diagram in urban road network under VANETs environment. To analyze the MFD under VANETs environment, a simulation test is conducted with grid road network in NS-3 simulation platform. The MFD of the simulated network is presented in the study. By comparing simulation results between different wireless communication scenarios, research shows that wireless communications under VANETs environment has an impact on the shape of the MFD for the simulated road network.

Future research will investigate the macroscopic fundamental diagrams under VANETs environment with field data and real urban network rather than grid network in simulation.

POSTER PRESENTATION

Real-Time Control of Queue Spillbacks on Signalized Arterials**MOHSEN RAMEZANI****NICOLAS DE LAMBERTERIE****NIKOLAS GEROLIMINIS***École Polytechnique Fédérale de Lausanne***ALEXANDER SKABARDONIS***University of California, Berkeley***INTRODUCTION**

In the last decades, drastically increased traffic demands in cities have resulted in congestion problems. The development of intelligent transportation system (ITS) technologies with new monitoring paradigms and computational tools enables to alleviate congestions by estimating traffic states in real-time and thus implementing traffic-responsive signal control schemes instead of fixed one, which is particularly not adapted to oversaturated conditions (1). Although an appropriate use of these new technologies will ameliorate traffic in local congested areas, an optimal traffic-responsive signal control in scale of an arterial or a network is computationally intractable.

In urban networks with signal-controlled intersections, when demand exceeds capacity queues fail to clear during the allocated green times that creates oversaturated traffic conditions. Spillovers occur when growing queues at the downstream link block the arrivals from the immediate upstream link such that vehicle queues cannot discharge at capacity, although the signal phase is green. Spillbacks may also occur when turning vehicles fill up the available storage length of turn bays and block the through movements. This results in a significant reduction of the intersection capacity and also excessive delay both for the through traffic movement and the cross-street traffic entering the arterial.

This paper advances the state-of-the-art of traffic signal control in congested urban arterials by integrating a clustering approach into a control policy of long queues prior to spillover occurrence to reduce the risk of spillovers through a feedback strategy. First, we introduce an arterial clustering approach that detects in real-time the links with long queues along one direction of the arterial, clustering them together if they are consecutive and then identifying the entrance and exit intersections of each cluster. These intersections indeed constitute critical junctions (considering the analogy of an active bottleneck in a freeway) and therefore, proper adjustment of the signal timing settings in those intersections could improve traffic conditions in the whole arterial. Thus, it enables to implement locally smaller-sized decentralized signal control strategies while ensuring at the same time the global coherence of these strategies along the arterial. In this manner, we seek to improve traffic on arterial at a very low cost by acting merely on critical intersections, as opposed for instance to network-wide optimization process. This approach is adaptive and does not require information about turning movements, which is difficult to be estimated in real-time.

Hence, the purpose of this paper is to develop an elegant signal control strategy based on the arterial clustering approach that enables to act only locally on specific intersections. Including an

advanced detection of oversaturated states and a specific focus on queue spillovers prevention, it leads to significant reduction of congestion and thus improves the network traffic conditions (2).

ARTERIAL PARTITIONING

The clustering approach consists in detecting in real-time the arterial links with long queues, grouping them together if they are adjacent, and consequently identifying the entrance and exit intersections of each cluster. These intersections constitute critical junctions along the arterial. Thus, by proper modifications of the signal settings at those locations one could expect improvements in the traffic situation of the whole arterial. The developed methodology can be integrated in existing coordinated or actuated strategies as it intervenes in only a few intersections, with the objective of decreasing the risk of spillbacks.

The arterial partitioning first utilizes a criterion to differentiate oversaturated or close-to-spillback links from the uncongested ones based on the queue length at the end of each red phase. It is assumed that queue lengths can be estimated with some error from loop detectors or probe vehicle data by using existing state-of-the-art techniques (3).

A cluster of congested links consists of consecutive congested links in the arterial delimited with noncongested links. The entrance intersection designates the signalized intersection located at the upstream limit of the group of congested links. This intersection is denoted by l and the link just downstream $l + 1$. Correspondingly, the exit intersection is located at the downstream limit of the group of congested links and is denoted by f as well as the link just upstream (see Figure 1a).

The choice of the threshold to decide whether a link is considered as congested or not is based on a preventive strategy that detects moderate queues that exceed the capacity of the corresponding approach at the downstream intersection. Hence, considering arterial partitioning for cycle $k + 1$, we propose: if $q_k^i > q_{cr}^i$, link i is considered as congested, where q_k^i is the queue (vehicles per lane) in link i at cycle k and $q_{cr}^i = g_{init}^i \cdot s$ which is maximum queue of vehicles per lane in link i that can completely be cleared during the next green phase g_{init}^i for the arterial through movement with fixed-time control (s denotes the saturation flow per lane).

The pseudo code for arterial partitioning approach is as follows:

1. *If* cycle k of all intersections has been completed
2. Loop $i = 1, 2, \dots, i_{max}$ ($i = 1$ stands for the most upstream link in the arterial)
3. *If* $q_k^i > q_{cr}^i$
 - If* ($i = 1$) or ($q_k^{i-1} \leq q_{cr}^{i-1}$)
 - Append* intersection $i - 1$ to the entrance intersections l
 - Else*
 - Append* intersection $i - 1$ to the interior intersections
 - End If*
 - If* ($i = i_{max}$) or ($q_k^{i+1} \leq q_{cr}^{i+1}$)
 - Append* intersection i to the exit intersections f
 - End If*
- End If*
- End Loop*

Figure 1 illustrates arterial partitioning, as the demand creates queues that spill back upstream San Pablo–University critical intersection on several links up to San Pablo–Dwight intersection with fixed-time signal control. Congestion is consistently present from cycle 25 until the end of the simulation when very low demand enables all long queues to clear. Hence, San Pablo–University intersection is an exit intersection for the larger part of the simulation period (see Figure 1c), whereas San Pablo intersections with Addison, Allston, Dwight, and Grayson are successively entrance intersections during the simulation. Arterial partitioning and fixed-time signal control in Figure 1c reveals that besides these queue spillovers phenomena, Cedar and Gilman streets also face oversaturated conditions (i.e., $q_k^i > q_{cr}^i$) for about half of the simulation. Thus, arterial partitioning methodology appears also to be a valuable performance measurement tool independently of signal controlling.

SIGNAL CONTROL STRATEGY

Arterial partitioning at each cycle identifies group of congested links that are delimited with exit and entrance intersections. The seminal intuition of the proposed signal control strategy is to act only on exit and entrance intersections along the arterial. If one considers a critical group of congested links as a whole, to regulate its accumulation to the uncongested state, we need to manipulate the inflow and outflow of the critical group of congested links. Thus, the general direction is to increase the green phase duration of the arterial through movement at the exit intersection (increase the outflow) and also, to decrease the corresponding green phase duration at the entrance intersection (decrease the inflow). The corollary of such control strategy is that the interior intersections of a group of congested link are still controlled with initial fixed timings.

Exit Intersection Signal Control

Arterial partitioning is triggered once the queue in a link exceeds the corresponding intersection's capacity. At cycle $k + 1$ for which exit intersection signal control is activated, green extension is based on the difference between previous queue length q_k^f and q_{cr}^f . In addition, when the risk of a potential queue spillover in link f becomes significant ($s_k^f < s_{cr}$) another compensatory term is added to the green extension. Hence, the control strategy for the first cycle of activation is:

$$g_{k+1}^f = g_{init}^f + K[q_k^f - q_{cr}^f] + K_{s1}[s_{cr} - \min(s_k^f, s_{cr})] \quad (1)$$

where g_{k+1}^f is the cycle $k + 1$ green time for the through arterial movement at intersection f , and K and K_{s1} represent time increments (in seconds) for each additional vehicle queuing, respectively, beyond q_{cr}^f and passed s_{cr} threshold. For the next cycles as long as $q_k^f > q_{cr}^f$, a feedback control is implemented based on the attributes of the two last cycles, k and $k-1$:

$$g_{k+1}^f = g_k^f + K[q_k^f - q_{k-1}^f] + K_{s1}[\min(s_{k-1}^f, s_{cr}) - \min(s_k^f, s_{cr})] \quad (2)$$

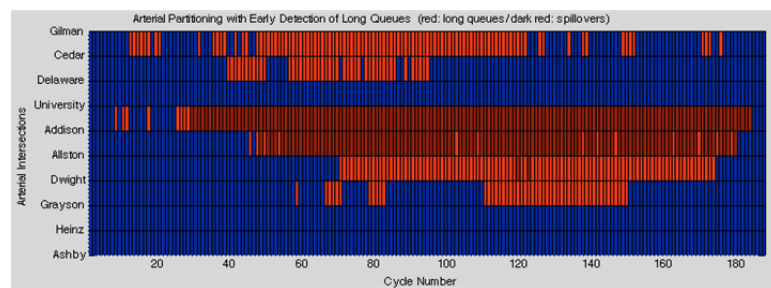
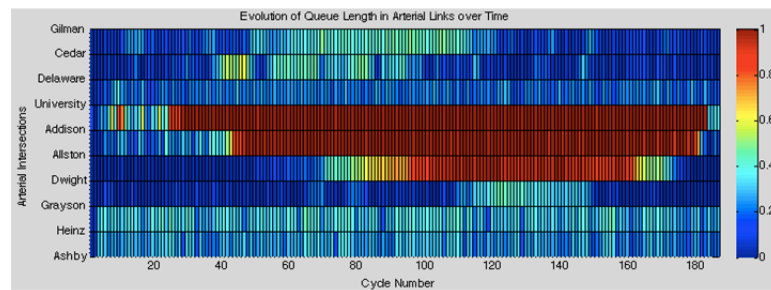
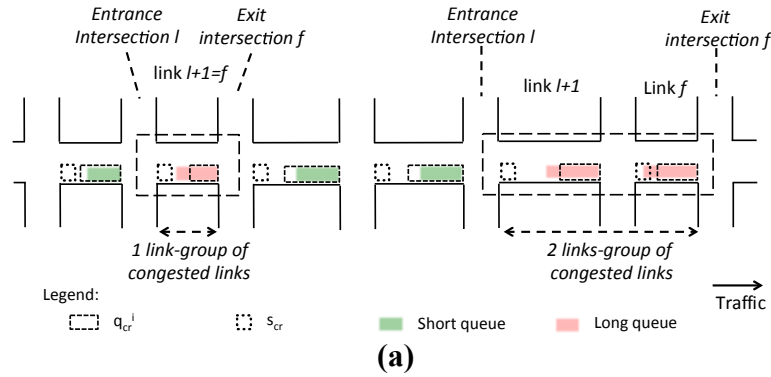


FIGURE 1 Arterial partitioning of San Pablo Avenue (Berkeley, California) northbound approach: (a) schematic example of arterial partitioning; (b) evolution of link queue lengths over time with fixed-time control (links queue occupancy from 0 to 1); and (c) arterial partitioning. (Note: blue = no detection of long queue; red = long queue detection; dark red = queue spillover detection).

Note that, feedback controller parameters, K and K_{s1} , have to be chosen according to proper control engineering methods or manual fine-tuning to maximize the reduction of the network delay time, however, feedback controllers are shown to be robust to a moderate range of parameter values (4).

Entrance Intersection Signal Control

The objective of the entrance control is to prevent or at least delay queue spillovers occurrence. Therefore, reduction of the arterial through movement green time at the entrance intersection is

only profitable when there is a substantial risk of queue spillovers occurrence, that is when $s_k^{l+1} < s_{cr}$. Hence, we propose the following feedback control:

$$g_{k+1}^l = g_k^l - K_{s2} [\min(s_{k-1}^{l+1}, s_{cr}) - \min(s_k^{l+1}, s_{cr})] \quad (3)$$

where K_{s2} is a parameter to be calibrated to maximize the reduction of the delay time.

REFERENCES

1. Geroliminis, N., and A. Skabardonis. Identification and Analysis of Queue Spillovers in City Street Networks. *IEEE Transactions on Intelligent Transportation Systems*, No. 99, 2011, pp. 1–9.
2. Skabardonis, A., and N. Geroliminis. Real-Time Monitoring and Control on Signalized Arterials. *Journal of Intelligent Transportation Systems*, Vol. 12, No. 2, 2008, pp. 64–74.
3. Ramezani, M., and N. Geroliminis, Exploiting Probe Data to Estimate the Queue Profile in Urban Networks. *Proc., 16th International IEEE Annual Conference on Intelligent Transportation Systems. ITSC*, 2013, pp. 1817–1822.
4. Keyvan-Ekbatani, M., A. Kouvelas, I. Papamichail, and M. Papageorgiou. Exploiting the Fundamental Diagram of Urban Networks for Feedback-Based Gating. *Transportation Research Part B*, Vol. 46, No. 10, 2012, pp. 1393–1403.

POSTER PRESENTATION

Macroscopic Evaluation of Incident-Induced Driver Behavior Changes

ZIAUR RAHMAN

STEPHEN P. MATTINGLY

University of Texas at Arlington

Reducing the impact of incidents on the freeway network, which represents a key goal for most regional traffic management systems, requires an effective traffic incident management system. Early and accurate recognition of incident impact the network represents one of the key challenges associated with incident management. Microscopic simulation represents an effective strategy for forecasting incident impact because it provides opportunities to study diverse conditions that may occur irregularly both spatially and temporally in the field. Often, the traffic simulation parameters appear be calibrated for uncongested conditions. This research explores macroscopic network performance when applying microscopic simulation models calibrated under both congested and uncongested conditions; the research considers a default uncongested parameter set and compares it with a parameter set calibrated using macroscopic data during congested conditions and a parameter set calibrated using microscopic data during congested conditions. Using a testbed along I-210 in Pasadena, California, the study varies incident duration, traffic volume, and lane closures to investigate the change in results across all three models. The study also investigates six actual incidents and compares the simulation results with the observed field conditions collected from the California Department of Transportation Performance Measurement System database. During the experimental cases, the macroscopic congested parameter set consistently estimates longer network travel times and slower speeds. The comparison with the field conditions confirms that the macroscopic congested parameter set better reflects field performance especially during more congested conditions. This research indicates the need to use congested conditions for calibration when investigating incident impact.

INTRODUCTION

An effective traffic incident management system has a crucial impact on reducing or minimizing the impact of incidents on the freeway network; it represents one of the key goals for most regional traffic management systems. The blockage or closure of freeway lanes causes significant delays on the road network (Kwon et al., 2006). These temporary blockages may cause changes in driving behavior in both blocked and unblocked lanes. As incidents create queues, delay, rubbernecking, secondary crashes, and many other problems, the accurate assessment of incident impacts appears critical for effective operational management policies.

Microscopic traffic simulation provides opportunities to study diverse experimental traffic conditions that may occur irregularly both spatially and temporally in the field. Often the parameters in the traffic simulation software appear to only be calibrated for the uncongested situation. Recent research (Gomes et al., 2004) indicates that microscopic traffic simulation parameters may actually change during incident conditions. Within the incident zone, which is where the lane numbers are reduced, Knoop et al. (2009) shows that the driver reaction time increases significantly and produces a 30% lower queue discharge rate at the macroscopic level. In this study, the researchers use traffic

simulation to investigate the impact of incident induced changes in microscopic behavior parameters' on macroscopic performance.

Woody (2006) finds that the VISSIM parameters—CC0, CC1, CC2, CC4, and CC5—have more influence on the capacity of a freeway section than other parameters. Sheu (2006) models intra-lane and inter-lane traffic behavior. Sheu calibrates the microscopic travel behavior characteristics: maximum acceleration, maximum deceleration, minimum safety spacing, and mean reaction time. The current research integrates these four traffic behavior parameters in a VISSIM model and compares the network performance resulting from an incident when VISSIM is calibrated based on congested and uncongested conditions. The definition of maximum acceleration from Sheu's research is similar to the definition of driver behavior parameter CC9, which is actually acceleration at 50 mph. According to PTV VISION, the maximum deceleration is the maximum technically feasible deceleration. During an incident induced situation, driver behavior impacts both car-following and lane changing based on the minimum safety spacing and mean reaction time. The minimum safety spacing is similar to standstill distance, and mean reaction time is compared with CC1 (headway time). The study investigates the following research questions:

- What are the differences in macroscopic performance measures (such as speed, flow, and delay) when congested freeway conditions are simulated using traffic behavior parameters calibrated under congested conditions versus those calibrated under uncongested conditions?
 - Do the number of lanes blocked during an incident affect the magnitude of these differences?
 - What effect, if any, does incident duration have on these differences?
 - Does the amount of demand affect these differences?
 - Do the uncongested-based parameters or the congested-based parameters more accurately reflect observed field performance during congested conditions?

METHODOLOGY

Test Environment

This paper considers a previously studied 15-mi long section of I-210 West between Vernon St. and the SR-134 junction in Pasadena, California (Figure 1). This site is selected because it has both California Department of Transportation (Caltrans) Performance Measurement System (PeMS) data (e.g., speed, flow, and occupancy) and California Highway Patrol incident data available. Furthermore, an earlier study specifically calibrates VISSIM for this site using qualitative operational criteria during congested conditions (Gomes et al., 2004). The current study uses PeMS data sources during both the experimental and validation phases of the project.

This freeway section has a high-occupancy vehicle (HOV) lane, a freeway connector, and 20 metered on-ramps with and without bypass lanes. A median side HOV lane spans the entire study network and is separated by intermittent barrier. The Gomes et al. (2004) test network uses these characteristics to replicate the exact field conditions. However, the current study only considers the main lane movement of eastbound and westbound I-210 because the study focuses on the car-following and lane-changing behavior specifically related to an incident rather than allowing those resulting from weaving, merging, and diverging to interfere. As a result, the study does not consider any HOV or on- and off-ramp movements. The research design treats HOV lanes as a regular lane

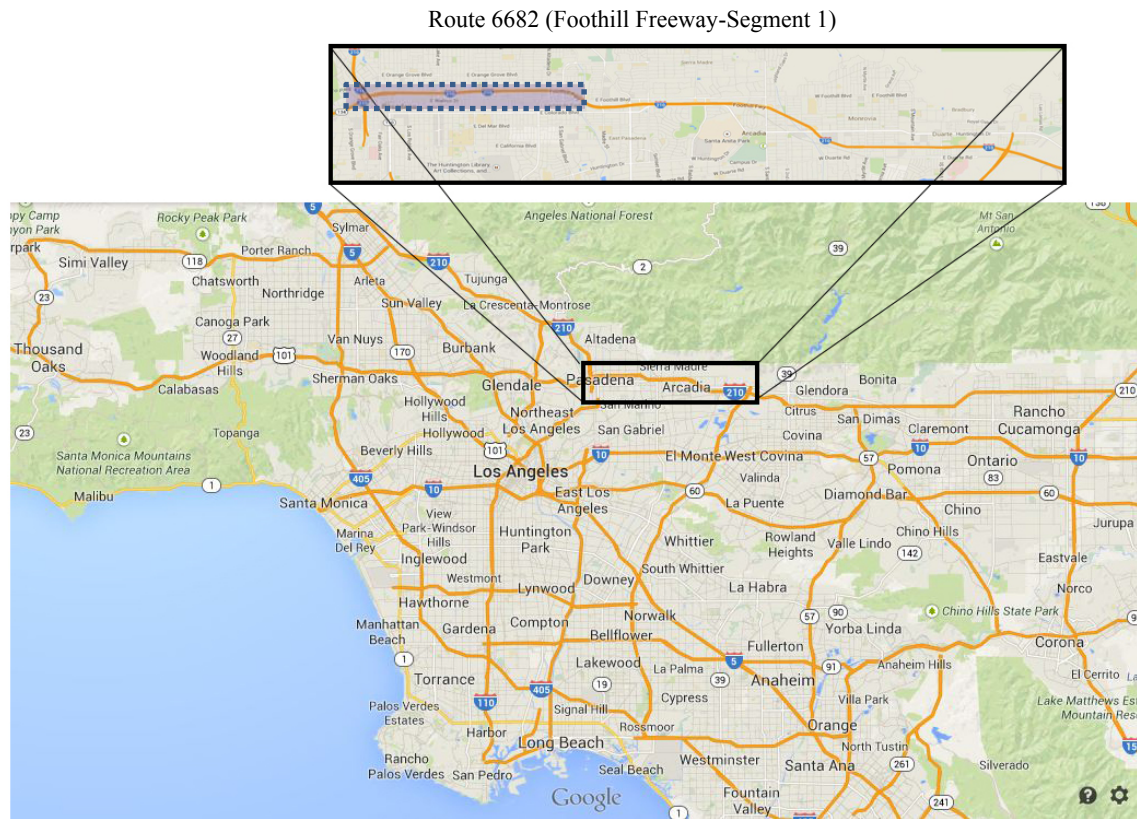


FIGURE 1 Geographic position of the study network (Google Map, accessed on July 2014).

with the main lane movement because the incident location of gives the drivers access to the HOV lane. Figure 1 shows the location of the study network in Los Angeles.






VISSIM simulations require a comprehensive and complete description of the network layout for a realistic output, and the recommended method for entering geometric information is to construct a scaled map, in a bitmap or jpeg format. Using 32 high-resolution Google pictures from Google map, the author makes a single panorama view of the network with the photo merge command of Adobe Photoshop. The research team uses the network graphic as the background for the VISSIM network. Important features for replicating the test bed include:

1. Number of lanes;
2. Width of lanes;
3. Lane drops;
4. Auxiliary lanes;
5. Location of the HOV lanes;
6. Location of data collection points; and
7. Location of the experimental incident.

The EB traffic enters the network at Orange Grove Boulevard and leaves the network at N. Vernon Avenue (Figure 1). Table 1 shows the layout of various network sections with lane configurations and milepost positions. Throughout the test bed, the number of lanes varies in different sections. Depending on the configuration of the lanes, an incident can have numerous effects on capacity. Where SR-134 starts to merge with I-210, the minimum number of lanes is

four plus a HOV lane. Vehicle Detection ID 717606 (Table 1) is situated at Orange Grove Avenue near milepost R13.18. This initial section is only 0.53 mi long and has four main lanes, one HOV with no access in this section, and no auxiliary lane. Similarly, the section that starts at N. Marengo Avenue is almost 1 mi long and has an auxiliary lane along with five main lanes and one HOV lane. Here, the HOV lane still has no entry or exit. But the last section of the study network near I-605 to I-210 off-ramp has a different configuration. The auxiliary lane here starts at the I-605 to I-210 off ramp and ends at N. Irwindale Avenue. The whole section has four main lanes and access between the main lanes and HOV lane. Throughout the study section, the lane width varies from 11 to 12 ft. The inner shoulder width varies from 2 to 12 ft whereas the outer shoulder width remains 10 ft in most sections. The experimental case considers an incident in the eastbound direction just after the N. Lake Avenue overpass, which has five main lanes and one HOV lane. The validation case uses different locations on the eastbound direction because they have available incident information. The research team places VISSIM queue counters, travel time sections, and data collection stations in these sections for performance measurement. Queue counters provide average queue length, maximum queue length, and number of vehicles stopped in the queue. A queue counter also measures the longest queue in the network. For both the experimental and validation cases, queue counters help identify the presence of any queue upstream of the incident during the simulation run and help calculate the queue dissipation time. A travel time section helps in the validation process where only Route 6682 of I-210 (Figure 1) is considered and it can easily measure the travel time from origin A to destination B. The data collection sections measure the microscopic level behavior such as speed, acceleration, and occupancy of each vehicle in the network.

TABLE 1 Layout of Study Network Sections

ID	Schematic	Start Name	CA PM	Abs. PM	Length	Lanes + HOV + Auxiliary	Auxiliary Start	Auxiliary End
717606		Orange Grove	R13.18	13.18	0.53	4+1+0	NA	NA
717631		Fair Oaks 1	R25.14	25.12	0.46	4+(1)+0	NA	NA
717633		Marengo	R25.74	25.72	0.68	5+1+1	N. Marengo Ave.	N. Lake Ave.
717646		San Gabriel	R28.7	28.68	0.91	5+1+1	N. Sunny Slope Ave.	N. Kinneloa Ave.
761206		Irwindale	R38.009	38.3	0.5	4+1/(1)+1	I-605 to I-210 E off-ramp	N. Irwindale Ave.

NOTE: ID = identification; CA = California; PM = post mile; abs. = absolute; NA = not available; N = north; ave. = avenue.

Traffic Demand

Instead of choosing an arbitrary traffic volume for the experiment, the researchers use the Caltrans PeMS database for setting likely high-volume cases where an incident will result in congestion. Relying on field data for traffic demand ensures that the network performance in response to an incident reflects real conditions while the focus on peak conditions increases the importance of the incident impacts. The study uses the hourly flow data from six randomly selected months in 2006 to identify the peak hour and peak flow rate for each day of the week. For the weekday cases, the peak hour occurs at either 8:00 a.m. or 5:00 p.m. For example, the most frequently occurring peak hour (mode) for Friday is 8:00 a.m. and the average peak hour flow rate for the eastbound direction is 5,486 vph (Table 2). This study considers both Tuesday and Friday peak demands to see if a moderate change in demand during the peak hour impacts the macroscopic performance measures during an incident. The following tables give both eastbound and westbound demands.

Incident Lane Blockage and Closure

Like traffic demand, the incident magnitude or number of affected lanes also impacts the congestion level. Due to lane drops and auxiliary lanes (Table 1), different sections of I-210 in the study test bed have different lane configurations; however, the minimum number of lanes is five. Throughout the experimental design, only main lane movements (four lanes) may be blocked by an incident while the HOV lane remains clear for all experimental trials. The researchers investigate the resulting impacts from nine different lane closure patterns. These include single lane closures for each of the four lanes and combination lane closures for

TABLE 2 Weekly Average Peak Hour Flow

Day	Time of Day	Average Hourly Flow (vph)
Eastbound Flow		
Friday	8	5,486
Saturday	12	4,917
Sunday	13	4,478
Monday	17	5,725
Tuesday	17	5,887
Wednesday	17	5,853
Thursday	17	5,756
Westbound Flow		
Friday	8	5,946
Saturday	12	5,715
Sunday	13	5,739
Monday	17	6,043
Tuesday	17	6,283
Wednesday	17	6,211
Thursday	17	6,186

contiguous lanes (e.g., lane 1 and 2, lane 2 and 3, lane 3 and 4, lanes 1 through 3, and lanes 2 through 4). Larger incidents require more lane changes and may cause the variation in performance between the different VISSIM models to increase. For validation purpose, three main lane movements and three HOV movements are considered and HOV lane movement is restricted for the HOV-class type vehicles that have a passenger occupancy of more than two.

Incident Duration

The duration of any type of incident also has an effect on the network performance because longer lasting incidents typically cause longer queues to develop. All of the experimental variables discussed to this point affect the incident’s impact on the network. The impact becomes more severe as the demand, the number of affected lanes and the incident duration increase. An initial review of the California PeMS data indicates that even short incidents typically last 30 min; therefore, the experiment considers two incident durations of 30 and 45 min representing relatively typical short incidents.

Figure 2 describes different combinations for the experimental variables in each simulation trial. Incident location is the highest level in this hierarchy. Two different traffic demands for Friday (5,486) and Tuesday (5,887) represent the next experimental dimension. The nine different lane closure combinations and two incident durations (30 and 45 min) represent the remaining experimental dimensions. In total, these create 36 different experimental trials.

Parameter Mapping

The parameter comparison (Table 3) provides evidence that the default VISSIM parameters in the uncongested condition are different from the I-210 West parameters, which have been calibrated based on congested conditions. After the test bed construction, the research team selects three different sets of microscopic driving behavior parameters for the experiment. The default VISSIM parameters represent the parameters calibrated under uncongested conditions. Two other sets of driving behavior parameters represent VISSIM calibrations under congested conditions (Sheu, 2013; Gomes et al., 2004). As an example of parameter mapping from earlier

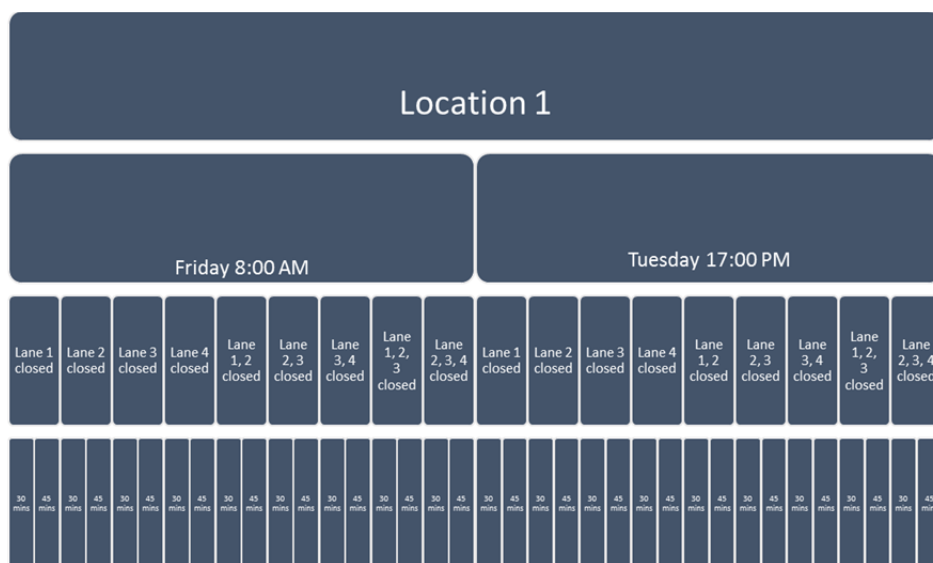


FIGURE 2 Variation of experimental combinations.

TABLE 3 Comparison of Calibrated Parameters

Driving Behavior Parameter Sets	Default VISSIM	Sheu (14)	I-210 (5)
Lane change (ft)			200–1,450
Maximum deceleration for own (ft/s ²)		-23.95	
Maximum deceleration for the trailing vehicle (ft/s ²)			
Waiting time before diffusion (s)			
Safety distance reduction factor			
CC0 (stand-still distance) (ft)	4.92	3.94	5.57
CC1 (headway time) (s)	0.9		0.9
CC2 (following variation) (ft)	13.12		13.12
CC3 (threshold for entering following)	-8		-8
CC4 (negative following threshold)	-0.35		-2
CC5 (positive following threshold)	0.35		2
CC6 (speed dependency of oscillation)	11.44		
CC7 (oscillation acceleration)	0.82		
CC8 (standstill acceleration)	11.48	11.68	
CC9 (acceleration at 50 mph)	4.92	7.28	

studies, the minimum safety distance, which Sheu (2013) describes in terms of vehicle length and stand-still distance, is related to the stand-still distance in the VISSIM model. The research team matches this and other similar parameters from the previous research with closely related VISSIM driver behavior parameters (those mentioned in Table 3) and generates the corresponding value for each of them. In VISSIM, different road classes have default parametric values. For the freeway section, only the highlighted parameters are changed for the Sheu and I-210 experimental models.

Experimental Design

The study uses three types (uncongested, congested–macro and incident–micro) of calibrated VISSIM models for the experimental test bed; for each type, variations are also evaluated based on the ranges of observed parameter values. The calibrated VISSIM model for I-210 is the base congested–macromodel while the default VISSIM model serves as the uncongested model. For each “experimental incident,” the study compares the three VISSIM model results. Figure 2 describes the experimental variables for each simulation trial, which is replicated for each VISSIM model. Section 3 discusses the measures included in this comparative analysis in detail.

Simulation Evaluation

This study uses two simulation output files, Data Collection and Network Performance. As its name indicates, the Network Performance output file describes network performance of the test bed. This file provides macroscopic parameters such as average speed, total delay, and total travel time. Queue counters and detectors are placed at various section of the test bed alongside with travel time sections. The queue counter output files provide the data for calculating the queue dissipation time. Travel time Station 16 is placed at the starting point of Route 6682 (Foothill Freeway, EB I-210-Segment 1). This station measures the time required for the vehicles to reach the end of this section.

MACROSCOPIC MEASURES OF EFFECTIVENESS

Each experimental case generates critical macroscopic performance measures like delay, travel time, and speed. These measures are used for comparative analysis and validation. The research investigates differences in macroscopic performance due to changes in microscopic travel behavior and the experimental variables. One measure, Travel Time Index, is the ratio of the actual average travel time to free-flow travel time. A lower index value shows better network performance and less incident impact.

Figure 3 compares four different combinations of average speed from the experiment. The left two graphs compare the speed for a 30-min incident for different driving behavior parameters. In both cases, the macroscopic congested parameters result in lower average speed. In most cases, Sheu’s microscopic congested parameters give a higher average speed than observed for the other cases. The same pattern is also found for the 45-min incident.

For the experimental incident location, the study also examines the travel time for the 30- and 45-min duration incidents with two different traffic demand and various lane closure combinations (Figure 4). As expected, the travel time increases with the increase in number of lanes affected by the incident; however, as incident impact increases (more lanes closed and longer time to clearance), the simulated models show greater variability. The travel time results reflect the speed results where the macroscopic parameters give a longer travel time and the microscopic parameters give a lower travel time. The network speed and total travel time charts show that the default parameters give a total travel time that lies between the I-210 and Sheu parameters. The



FIGURE 3 Average speed for different lane closure combinations and incident duration.

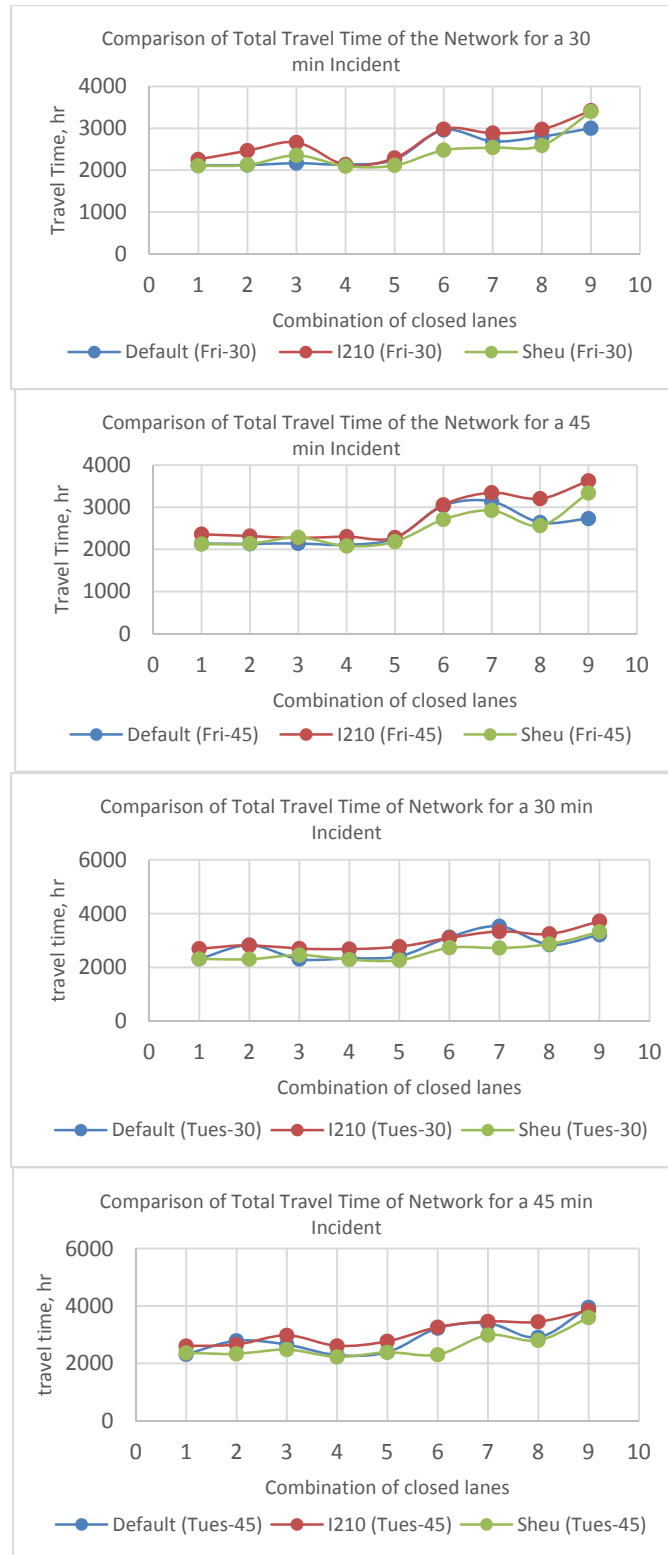


FIGURE 4 Total travel time for lane closure combinations and incident duration.

reason behind this could be the more accurate replication of the situation in the Gomes et al. model because in the general case an incident induced calibrated parameter should result in lower speed than the default parameters. The comparison of travel time for a 30-min-long incident (Figure 5) indicates that the incident duration has a significant impact on network performance.

The percentage of difference is the ratio of the difference (non-zero) of the default value and the two congested parameter cases. The results in Table 4 show that Sheu’s calibrated parameters have a smaller difference from the default for speed. For the single-lane closures, all of the differences for the Sheu model remain 8% or lower. In contrast, the I-210 model has differences in excess of 20% and over half of the single-lane closure experiments result in differences in excess of 10%. Sheu’s model departs further from the baseline model when more lanes are closed and longer incidents occur while the I-210 model does not appear to show the same if any change in speed difference when the incident increases in size.

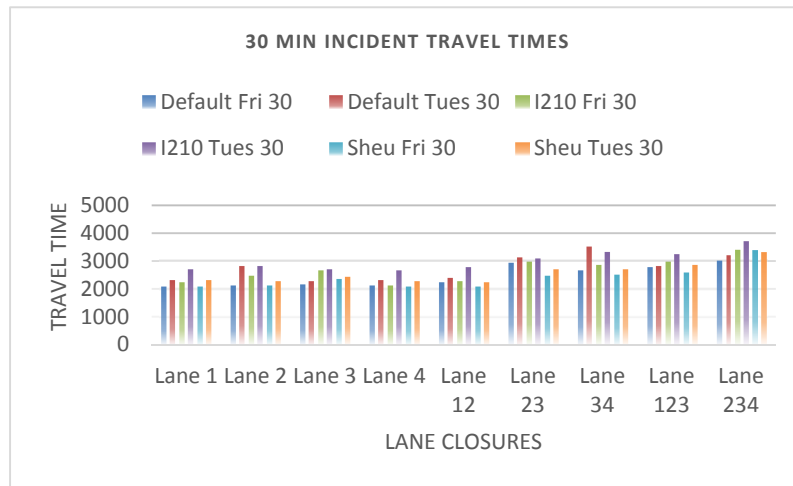


FIGURE 5 Travel time comparison for a 30-min incident.

TABLE 4 Percentage Difference in Travel Time Measurement

Calibration	Demand	Incident Length	Lane 1	Lane 2	Lane 3	Lane 4	Lane 12	Lane 23	Lane 34	Lane 123	Lane 234	
Absolute Percent error	I210	Friday	30	6.9%	16.1%	22.8%	0.2%	1.6%	0.7%	7.3%	6.1%	13.9%
		Friday	45	10.4%	8.6%	6.1%	9.2%	0.6%	0.7%	6.6%	21.3%	32.8%
		Tuesday	30	16.2%	0.1%	17.0%	14.8%	14.8%	0.9%	5.4%	14.1%	15.6%
		Tuesday	45	12.5%	4.7%	12.2%	14.1%	15.8%	1.1%	2.0%	18.4%	2.5%
	Sheu	Friday	30	0.3%	0.1%	8.6%	1.8%	6.2%	16.3%	5.6%	7.7%	13.1%
		Friday	45	0.7%	0.2%	6.8%	1.3%	3.7%	10.8%	6.8%	2.9%	22.3%
		Tuesday	30	0.1%	18.4%	6.5%	1.9%	6.2%	12.6%	22.7%	0.8%	3.2%
		Tuesday	45	2.2%	16.3%	6.5%	2.8%	0.2%	28.5%	12.0%	3.8%	9.0%

VALIDATION

Approach

The validation procedure compares the simulation outcomes with the field values of network performance during actual incidents on a selected set of measures of effectiveness. The observed incident impact can be assessed using macroscopic Caltrans PeMS data to generate the observed measures of effectiveness.

Failed detectors represent a common problem that research studies face while using detector data. Large quantities of data is readily available from California PeMS and the reliability of these data is pretty high. For validation, the reliability of the vehicle detector stations (VDS) remain above 85% for the three different incident days (Table 6), which indicates that the field results remain relatively valid.

The simulation results can be compared to the field data during the observed incidents. For each selected performance measure, an error indicator evaluates the relationship between the simulation output and the field data. Root mean square error (RMSE) measures this divergence.

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

where

y_i = is the observed value of the i th observation and
 \hat{y}_i = is the predicted value of the i th observation.

Considering the field observation as the observed value, the RMSE can be used as a measure of the spread of the y -values. Mean absolute percentage errors (MAPEs) can be used to study simulated traffic arrivals, average arrival speed, and lane changing fractions. The MAPE is defined as follows:

$$MAPE = \frac{100}{N} * \sum_{i=1}^N \left| \frac{x_i - \hat{x}_i}{x_i} \right| \tag{2}$$

where

x_i = is the field observation,
 \hat{x}_i = is the modeled or simulated observation, and
 N = is the number of data points.

TABLE 6 Percent of the PeMS Detectors Working

Detectors Working			
	June 1, 2014	June 9, 2014	June 30, 2014
Working	87.1	95.16	93.55
Not working	12.9	4.84	6.45
Total	100	100	100

Validation Sub-Network and Incident Description

Travel time along a segment of the network can be accumulated from the California PeMS; PeMS designates a segment of I-210 as Route 6682 (4.568 mi) and tracks travel time for this segment. The absolute start post mile (PM) of Route 6682 is 24.980 and the end PM is 29.548. For this route, the study collects field data from three vehicle detection stations (Figure 6). These stations are selected arbitrarily on the I-210 eastbound corridor (Route 6682–Foothill Freeway, eastbound, Segment 1) where incidents have occurred in real life. One downstream station captures the traffic conditions beyond the incident, which shows the number of vehicles departing the incident area. Another one is less than 1 mi upstream and captures the nearby incident impact. The final one is located almost at the beginning of the study section to track the number of vehicles entering the study section and calculate freeway demand; it should be in a location where the incident queue will not reach it. The study collects PeMS flow and speed measurements for these three locations every 5 min after the incident occurs.

The validation process considered six recent incidents—three HOV and three main lane—within Route 6682. The first incident, a stalled vehicle, blocked the HOV lane at Altadena (I-210–east PM 28.40) on Friday, May 2, 2014, at 14:48. After being assigned at 14:57, the California Highway Patrol (CHP) cleared the incident in 26 min. The second HOV incident, tire debris, occurred at the same location on Friday, May 16, 2014, at 15:55 p.m. and lasted about 12 min. The last HOV incident, a metal basket in the lane, occurred near Altadena (I-210–east PM 29.00) on Friday, June 6, 2014, from 11:30 to 11:42 a.m. None of the HOV incidents involved a crash; however, they could still have an impact on network performance.

The remaining three incidents occurred on the main lanes. The first incident, a stalled car, occurred in lane one on Sunday, June 1, 2014 at 14:30 near I-210–east PM 28.30 and required almost 6 min to clear. The second incident, tire debris blocking lane 3, took place at I-210–east PM 28.40 from 6:00 to 6:26 a.m. on Monday, June 9, 2014. The last incident, plywood blocking lane three, occurred at I-210–east PM 28.40 from 11:08 to 11:16 a.m. on Monday, June 30, 2014.

While none of these incidents involve crashes, they represent typical incidents that may regularly occur on any freeway. If a congested calibration reflects observed field conditions more closely for these incidents, larger incidents appear likely to magnify this importance. Table 7 summarizes the validation incident time and location data.

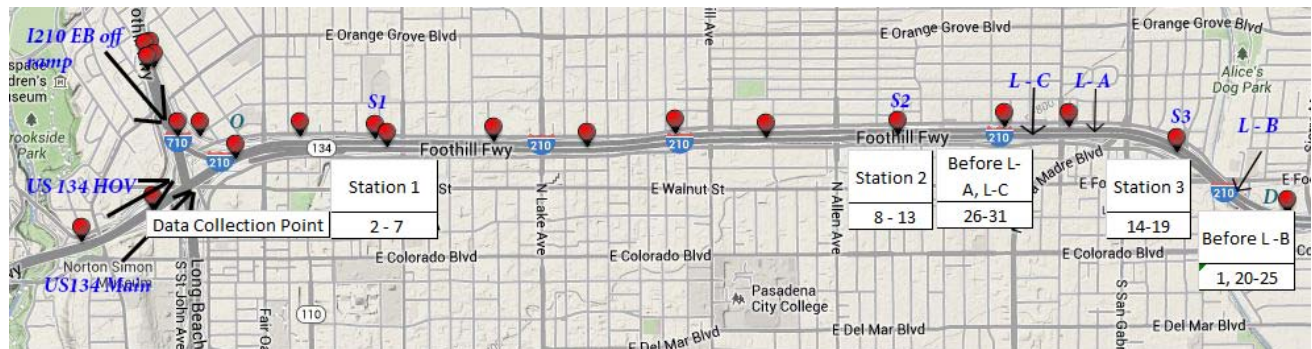


FIGURE 6 VDS position for data validation cases.

TABLE 7 Incident Location, Time and Duration

Date	Time of Occurrence	Duration (min)	Location	Travel Time (min)
5/2/2014	2:48:00 p.m.	26	HOV	14.3356
5/16/2014	3:55:00 p.m.	12	HOV	19.6648
6/1/2014	2:30:00 p.m.	6	Lane 1	4.0996
6/6/2014	11:30:00 a.m.	12	HOV	4.4616
6/9/2014	6:00:00 a.m.	26	Lane 3	4.0268
6/30/2014	11:08:00 a.m.	8	Lane 3	4.2428

The demand varies during each of these incidents, and PeMS data, shown in [Table 8](#), reflects this demand. Vehicles enter the subnetwork in three locations. Eastbound I-210 ramp enters from South Bound I-210 while the US-134 main lane movement occurs at the start of Route 6682. The US-134 HOV demand is either calculated from the US-134 main lane movement or collected from PeMS. The American Community Survey (2012) identifies that 11.09% of California workers over age 16 use the HOV lane. Due to the lack of HOV PeMS data, the validation sets the HOV demand for the HOV incidents at 11% of main lane demand.

Results

The research team creates travel time sections along route 6682 (starts at O and ends at D on [Figure 6](#)) so that the simulation travel times may be directly compared with the PeMS travel times. The simulated subnetwork travel times and field data appear in [Table 9](#). For the three main lane incidents, which do not appear to have much impact on traffic conditions based on the small changes in field travel times, the three parameter sets exhibit similar performance; the simulated differences remain 6% or less. The percent error from the field conditions remains thirteen percent or less, too. The HOV incidents, which have a more pronounced impact on field conditions, see a significant difference in the simulated results. For the HOV incidents, the macroscopic congested parameters consistently estimate longer travel times which more closely reflect field conditions. While the simulated results fail to capture the full impact of the first two HOV incidents, the macroscopic congested parameters recognize that significant congestion likely occurs. For the same cases, the other parameter sets reflect travel times at least 16% lower than I-210 parameters these differences increase to as much as 58% for the second incident. The differences with the field conditions appear even more pronounced they range from 19% to 80%. The validation shows that a model calibrated to capture congested conditions will produce a more accurate representation of network performance during incident conditions.

TABLE 8 Validation Demand

Location	HOV 5/2/2014	HOV 5/16/2014	HOV 6/6/2014	Main 6/1/2014	Main 6/9/2014	Main 6/30/2014
EB I-210 ramp	1,320	1,380	1,080	603	150	575
US-134 HOV	560	589	536	612	120	420
US-134 main	5,088	5,357	4,872	4,500	3,000	4,800

TABLE 9 Travel Time Validation

	HOV 5/2/2014	HOV 5/16/2014	HOV 6/6/2014	Main 6/1/2014	Main 6/9/2014	Main 6/30/2014
Field TT (min)	14.336	19.665	4.462	4.100	4.027	4.243
Default (min)	3.766	3.808	3.614	3.724	3.516	4.088
I-210 (min)	8.387	9.117	4.309	3.722	3.562	4.058
Sheu (min)	4.089	3.969	3.632	3.686	3.518	3.843

TABLE 10 Simulated versus Observed RMSE and MAPE

Simulation Parameter Set	RMSE	MAPE
Default	0.535	11.1
I-210	0.323	7.13
Sheu	0.565	12.7

The RMSE and MAPE (Table 10) reflect similar conclusions. The macroscopic congested parameters have a lower RMSE and MAPE than the other two parameter sets. The congested model appears much more capable of capturing the true impacts of congested conditions than the other two models. Sheu's model reflects the largest error because it estimates faster speeds as the typical conditions.

CONCLUSIONS AND RECOMMENDATIONS

This study examines simulation results for three different parameter sets (uncongested, macroscopic congested, and microscopic congested) during incident conditions. When comparing the results during the experimental incidents, the macroscopic congested parameter set consistently predicts longer travel times and lower speeds. The differences between the model results appear to increase as the incident magnitude increases (i.e., more lanes closed and longer clearance time); the microscopic congested parameter set exhibits a greater departure from the uncongested results as the magnitude increases. Due to the high volumes used in this study, the incident duration and number of lanes closed have a significant impact on network performance. The results indicate that the model likely has a significant impact on network performance during incident conditions and these differences become more important for larger incidents.

The second part of the study compares the same microscopic simulation parameter sets with observed field conditions inside the study network during actual incidents. For this case, the macroscopic congested parameters reflect field conditions much closer than the other parameter sets during significantly congested conditions and generate similar results with the other models during the relatively uncongested incidents. While the macroscopic congested parameters outperform the other parameter sets during congested conditions, they still fail to adequately capture incident impacts; this indicates the need to consider congested lane changing and lateral spacing behavior to capture the full incident impact. Lane-changing behavior can significantly change the queue accumulation in a heavily congested network and hence additional parameter

mapping appears necessary. Incident conditions also introduce the concept of compulsory lane changing maneuvers where vehicles upstream of an incident site must vacate all blocked lanes. The macroscopic parameters have a larger standstill distance, which may reflect Knoop et al.'s findings that the driver reaction time increases during an incident. Curiously, the microscopic congested parameter set shows the greatest departure from field conditions, however, this may reflect the challenge of mapping one microscopic model into another model rather than its actual performance. This issue needs to be explored more closely by using microscopic data to directly calibrate VISSIM parameters. The study shows that a model calibrated to capture congested conditions produces a more accurate representation of network performance during incident conditions and it likely suffers little degradation in performance during uncongested conditions.

Additional future research should explore if the increase in reaction times near an incident translates to the opposing flow direction, and identify the parameters necessary to adequately capture the rubbernecking phenomenon. An ideal experiment will explore numerous sites where both site specific congested and uncongested calibration occurs; furthermore, an increase in the field cases to cover a wider range of lane closures, incident durations and demands appears valuable, too. For validation purposes, incorporating the observed origin–destination pattern and all network components appears critical because they can help mitigate the incident impacts; these rely on the location of the HOV's intermittent barrier and on- and off-ramps.

REFERENCES

- Kwon, J., M. Mauch, and P. Varaiya. The Components of Congestion: Delay from Incidents, Special Events, Lane Closures, Weather, Potential Ramp Metering Gain, and Excess Demand. Presented at 85th Annual Meeting of the Transportation Research Board, Washington, 2006.
- Gomes, G., A. May, and R. Horowitz. Calibration of VISSIM for a Congested Freeway. California PATH Program, Institute of Transportation Studies, California, 2004.
- Knoop, V. L., H. J. Van Zuylen, and S. P. Hoogendoorn. Microscopic Traffic Behaviour Near Incidents. *Proc., Second Sino-Dutch Joint Workshop in Transportation and Traffic Study*, Shanghai, China, 2009.
- Woody, T. Calibrating Freeway Simulation Models in VISSIM. Master's thesis. 2006. Available at http://courses.washington.edu/cee500/VISSIMCalibration_FinalReport.doc.
- Sheu, J. B. Microscopic Traffic Behavior Modelling and Simulation for Lane-Blocking Arterials Incidents. *Transportmetrica A: Transport Science*, Vol. 9, No. 4, 2013, pp. 335–357.
- Fellendorf, M., and P. Vortisch. Microscopic Traffic Flow Simulator VISSIM. *Fundamentals of Traffic Simulation*, International series in Operational Research and Management Science 145, DOI 10.1007/978-1-4419-6142-6_2, 2010.
- Mai et al. Protocol for VISSIM Simulation. Oregon Department of Transportation, 2011. Available at www.oregon.gov/ODOT/TD/TP/APM/AddC.pdf.
- Santhanam, S., and P. Byubgkyu. Development of VISSIM Base Model for Northern Virginia (NOVA) Freeway Systems. Research Report No. UVACTS-13-0-124. Center for Transportation Studies at the University of Virginia, 2008.
- Panwai, S., and D. Hussein. Comparative Evaluation of Microscopic Car-Following Behavior. *IEEE Transactions on Intelligent Transportation System*, Vol. 6, No. 3, 2005, pp. 314–325.
- Means of Transportation to Work by Selected Characteristics for Workplace Geography. 2012 American Community Survey 1-Year Estimates. S0804. U.S. Census Bureau. Available at http://factfinder2.census.gov/rest/dnldController/deliver?_ts=425214164750.

- Gipps, P. G. A Model for the Structure of Lane-Changing Decisions. *Transportation Research Board*, Vol. 20B, No. 5, 1986, pp. 403–414.
- Hidas, P. Modelling Lane Changing and Merging in Microscopic Traffic Simulations. *Transportation Research Part C*, Vol. 10, 2002, pp. 351–371.
- Google Map. <https://www.google.com/maps/@34.142023,-118.0593021,9680m/data=!3m1!1e3>.
- Hidas, P. Modelling Vehicle Interactions in Microscopic Simulation of Merging and Weaving. *Transportation Research Part C*, Vol. 13, 2005, pp. 37–62.

POSTER PRESENTATION

Current Validation of Traffic Models *Are They Correct for Uncertainty Analysis?*

JORDI CASAS

JOSEP PERARNAU

TSS-Transport Simulation Systems, Spain

VINCENZO PUNZO

MARCELLO MONTANINO

Università di Napoli Federico II, Italy

From a methodological point of view it is widely accepted that simulation is a useful technique to provide an experimental test bed to compare alternate system designs, replacing the experiments on the physical system by experiments on its formal representation in a computer in terms of a simulation model. Simulation may thus be seen as a sampling experiment on the real system through its model. The reliability of this decision-making process depends on the ability to produce a simulation model representing the system behavior closely enough for the purpose of using the model as a substitute of the actual system for experimental purposes. This reliability is established in terms of validation of the model. Model validation is inherently a statistical process in which the uncertainty due to data and model errors should be account for. This paper review the current practice of traffic modeling validation methodology proposed by traffic simulation guidelines published by public administrations or organizations (FHWA, UK Higways, TfL, etc.) and compares the statistic techniques used in other simulation fields. On the other hand, the paper describes techniques associated to uncertainty analysis and sensitivity analysis for understanding the nature of each input parameter of the traffic simulation model and their contribution in the simulation outputs. The set of parameters of a traffic simulation could be partitioned into two independent sets, parameters related to the supply (topological information, speed limits, traffic control, etc) and demand (vehicle types and origin–destination matrices). This paper is focused on apply the uncertainty analysis on the demand side because it is an unobservable parameter and normally is estimated using undirected methods.

All cases are illustrated numerically with examples from real-life simulation projects with the Aimsun simulator.

POSTER PRESENTATION

A Time-Series Analysis of Highway Capacity
*Case Study of Georgia 400***SHANGJIA DONG****HAIZHONG WANG***Oregon State University***JIA LI***University of Texas at Austin***INTRODUCTION**

Capacity has been understood as a deterministic traffic volume threshold, above which the flow breaks down into queued or bottleneck conditions with heavy congestion while below which traffic will flow (1). This indicates if the capacity is 3,000 vph, the traffic flows when velocity is 2,999 vph, and it is congested when its 3,001 vph. However, empirical capacity observations suggest stochastic capacity variations over time which is hardly described in a deterministic manner. Elefteriadou et al. (2) showed the breakdown does not necessarily occurs at the same demand level. Minderhoud et al. (3) investigated the empirical capacity estimation method for uninterrupted flow and recommended the product limit method. Brilon et al. (4) extended this method and suggested a Weibull distribution to describe the stochastic capacity based on empirical observations from Germany. In addition, they applied the stochastic capacity concept to capture travel-time reliabilities in freeway networks. This stochastic capacity concept differentiated from the conventional fixed-capacity definition is a noticeable advancement by enabling a probabilistic nature of road capacities, however this concept still bears the flow of its inability to capture how road capacities vary over time (i.e., defining freeway capacity as a time series). As a result of the stochastic capacity concept, the travel time will be stochastic. Therefore, the significant impacts of stochastic capacity on travel time reliability can be conducted.

OBJECTIVE

The purpose of this research is to provide a stochastic analysis of highway capacities using empirical data from Georgia SR-400 (GA400) as evidences. This problem is an intricate one as highway capacity depends on diverse dynamic factors, such as weather conditions, vehicle composition, heterogeneous driver behavior and static factors, including road geometric conditions, vehicle characteristics. Accurate modeling of capacity variation has to take these numerous factors into consideration, which may require substantial empirical data to unveil the underlying capacity variation patterns. Since the capacity changes day by day and form a series according to the time, we can look the capacity changing from a different perspective, namely,

the time series. In this paper, a time series method for modeling the capacity series will be developed, also, the suggested approach with a case study and experimental results would be illustrated.

MOTIVATION

The impetus to model highway capacity as a stochastic process is largely motivated by the fact that empirical highway capacity observations show stochastic variations over an extended observation duration as is verified from [Figure 1](#). A stochastic analysis of highway capacity is essential to understand the freeway bottleneck breakdown, congestion dynamics, and prediction of travel time reliability (5). Capacity is typically defined as the maximum amount or number that can be continued or accommodated. According to the current published version of the *Highway Capacity Manual 2010* (HCM 2010), the capacity of a system element is defined as “the maximum sustainable hourly flow rate at which persons or vehicles reasonably can be expected to traverse a point or a uniform section of a lane or roadway during a given time period, under prevailing roadway, environmental, traffic and control conditions. Highway capacity can be expressed either in vehicle or person. Without further clarification, the capacity here is vehicle capacity by default unless otherwise specified.

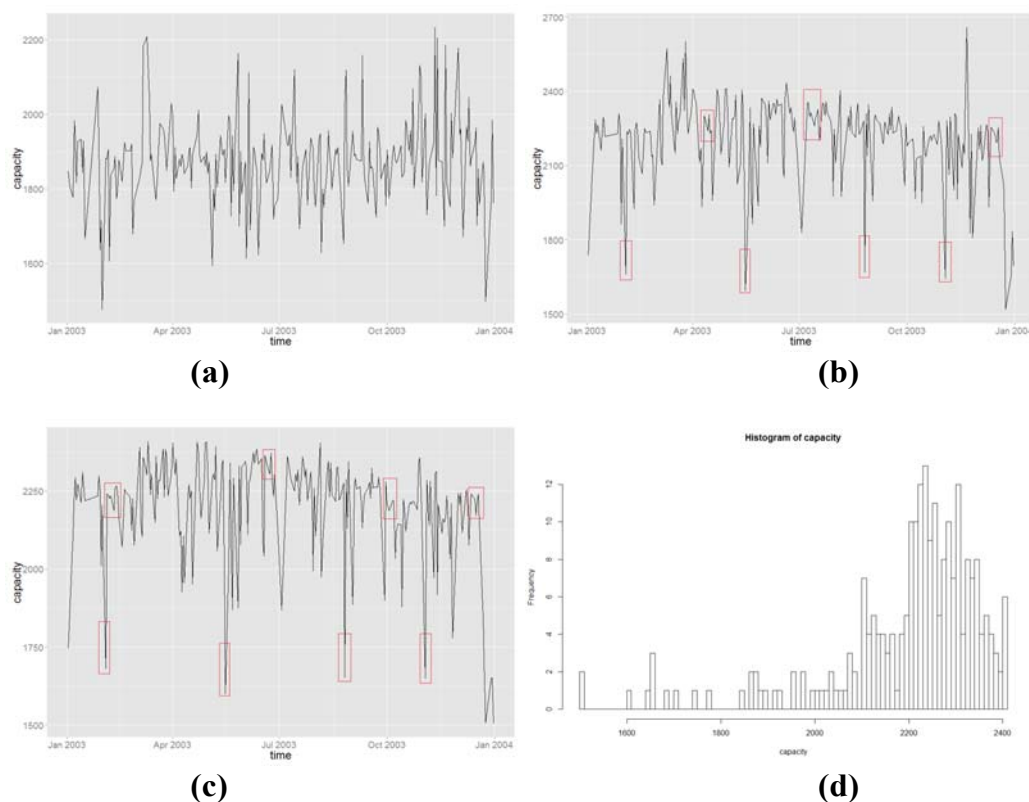


FIGURE 1 Capacity variation patterns on basic freeway segment from multiple GA400 detectors over a whole year's weekday observations: (a) 4000040, (b) 4001119, (c) 4001118, and (d) 4001118.

EMPIRICAL CAPACITY

Estimation of Empirical Capacities

According to HCM 2010, capacity is the maximum hourly rate at which vehicles pass a point at a given time period. When estimating capacity from empirical data, the 15-min flow rate is usually used. In the original GA400 dataset, traffic was reported every 20 s regarding the vehicle counts (i.e., cars, vans, and trucks). A passenger-car equivalent factor is used to convert the number of heavy vehicles to the equivalent number of passenger cars. The 15-min flow rate is scaled up to an hourly rate by multiplying 4, and the capacity of this location is estimated by the maximum hourly rate of each day. Repeat this process for every single weekday over the year 2003, a series of highway capacities and its variations over time is therefore obtained and this capacity variation trend can be described by a time series analysis. Next, the empirical capacity variations are presented at the freeway basic segment and the freeway on/off-ramps.

Capacity Variations on Basic Freeway Segment

As shown by Figure 1, the capacity does fluctuate on a daily basis within an interval: approximately 1,600 to 2,400 vphpl for freeway basic segment, and from 250 to 750 vphpl for freeway on/off-ramps. It is meaningful to describe the how capacities are distributed as is shown in Figure 1*d*. A Weibull distribution has been suggested and used in previous studies to describe it, however the statistical distribution fails to capture the capacity variations over time if there are interests to predict the future capacity changes. Therefore, a time-series analysis is adopted to bridge this gap which is appealing for prediction purposes.

Capacity Variations at On- and Off-Ramps

The capacity variation pattern at the freeway on/off-ramps is explored to reveal that this capacity variation pattern is not unique to basic freeway segments. The pattern is observed and verified in Figure 2. Therefore, the modeling and analysis methodology for a stochastic analysis of capacity variations on freeway basic segments is transferable to on- and off-ramps.

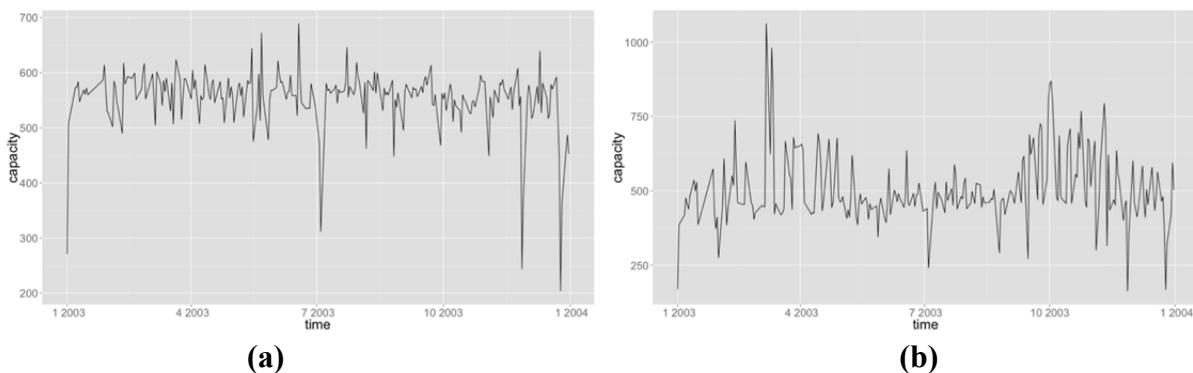


FIGURE 2 Capacity variations at on- and off-ramps from GA400 over a year's weekday observations: (a) on-ramp and (b) off-ramp.

PRELIMINARY RESULTS

Initial Spectral Analysis

A case study is designed to demonstrate the methodology using the GA400 dataset. Since the auto-correlation function (ACF) and partial auto-correlation function (PACF) did not suggest an obvious fit to a certain model. To investigate seasonality over the whole year 2003, a spectral analysis is employed to explore the trend in the capacity series to assess the seasonality characteristics of how capacity essentially varies over time for all the weekdays as is shown in Figure 3. The Figure 3a shows a smoothed spectrum from the capacity series estimated from one detector on GA400. In order to show how the seasonality is somehow related to how capacity changes on a particular weekday (i.e., Monday, Tuesday, Wednesday, Thursday, and Friday), the individual capacity series are plotted from Monday to Friday as is shown in Figure 3.

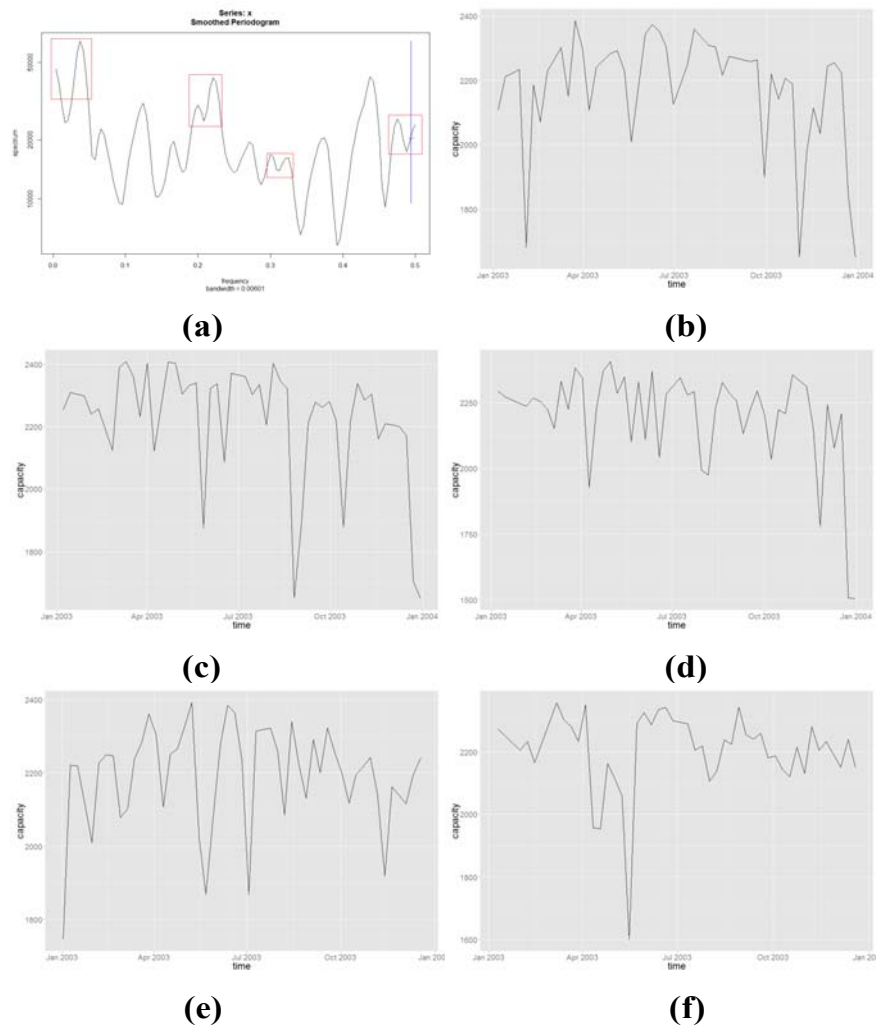


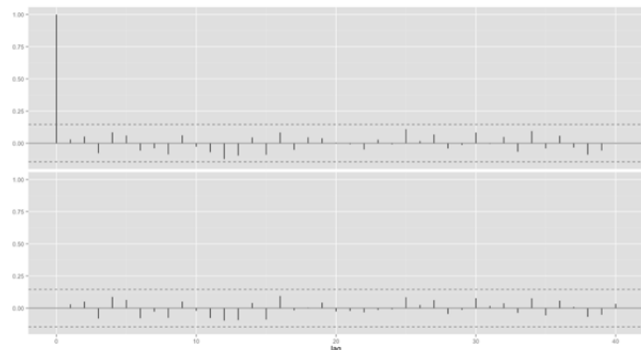
FIGURE 3 Spectral analysis and Monday through Friday capacity series: (a) spectral, (b) Monday, (c) Tuesday, (d) Wednesday, (e) Thursday, and (f) Friday.

Initial Time Series Analysis

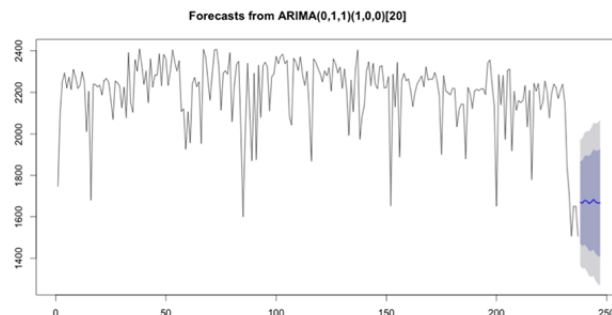
After conducting the spectral analysis, there are peaks shown on the periodogram, it reaches the highest point at $\omega = 0.05$. Let $T = 2\pi/\omega$, one gets $T \approx 20$, which is approximately a month excluding the weekends. (The data set is weekday data, so around 20 days one month.) Fit the capacity variations with an ARIMA $(0, 1, 1) \times (1, 0, 0)_{20}$ model, the AIC indicates the lowest value, which seems suggesting a monthly seasonality on the capacity variation. In addition, residual is fairly close to the white noise, which shows the proposed model describes the original data faithfully. As a result, a short-term forecasting of highway capacities can be performed to predict future capacity changes (Figure 4).

CONCLUSION

A preliminary stochastic time series analysis of highway capacities is provided and verified from the empirical observations both from basic freeway segments and on/off-ramps of Georgia SR-400. Modeling capacities as a stochastic time series instead of a statistical distribution function has obvious benefits to traffic operations analysis by knowing the temporal dimension of how capacity changed over time (existing) and how capacity might change in the near future (i.e., short-term prediction).



(a)



(b)

FIGURE 4 Model residual's (a) ACF and PACF and (b) 10 days short-term forecasting.

REFERENCES

1. Elefteriadou, L., and P. Lertworawanich. Defining, Measuring, and Estimating Freeway Capacity. Presented at 82nd Annual Meeting of the Transportation Research Board, Washington, D.C., 2003.
2. Elefteriadou, L., R. P. Roess, and W. R. McShane. Probabilistic Nature of Breakdown at Freeway Merge Junctions. In *Transportation Research Record 1484*, TRB, National Research Council, Washington, D.C., 1995, pp. 80–89.
3. Minderhoud, M. M., H. E. I. N. Botma, and P.H. L. Bovy. Roadway Capacity Estimation Methods Explained and Assessed. In *Transportation Research Record 1572*, TRB, National Research Council, Washington, D.C., 1997, pp. 59–67.
4. Brilon, W., J. Geistefeldt, and M. Regler. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. *Proc., 16th International Symposium on Transportation and Traffic Theory*, Vol. 125143, 2005.
5. Jia, A. Stochastic Capacity at Freeway Bottlenecks with Application to Travel Time Prediction. Ph.d. dissertation. North Carolina State University, 2013.

POSTER PRESENTATION

Collaborative Merging Behaviors and Their Impacts on Freeway Ramp Operations Under Connected Vehicle Environment

YUANCHANG XIE

HUIXING ZHANG

NATHAN GARTNER

TUGBA ARSAVA

University of Massachusetts, Lowell

Under connected vehicle environment, vehicles are able to communicate and exchange detailed information such as speed, acceleration, and position in real time. Such information is important for improving traffic safety. In the meantime, it allows vehicles to collaborate with each other, which may significantly improve traffic operations particularly at intersections and freeway ramps. To assess the potential benefits of collaborative driving behaviors enabled by connected vehicle technologies, this research proposes an optimization-based ramp control strategy and develops a simulation platform using VISSIM, MATLAB, and the Car2X module in VISSIM. In addition to the optimal control strategy, an empirical gradual speed limit control strategy is also proposed. These strategies are evaluated using the developed simulation platform in terms of average speed, average delay time, and throughput and are compared with a benchmark case with no control. The study results indicate that the proposed optimal control strategy can effectively coordinate all merging vehicles at freeway on-ramps and substantially improve safety and efficiency, especially when the freeway traffic is not oversaturated.

INTRODUCTION

In freeway ramp areas, frequent lane-changing and merging maneuvers can significantly reduce capacity and cause traffic congestion. These maneuvers may also jeopardize traffic safety, especially when vehicles from on-ramps have to first decelerate to a low speed due to congestion or lack of safe gaps, merge onto the freeway, and accelerate to normal speeds. Under connected vehicle environment, vehicles are able to communicate and exchange detailed information such as speed, acceleration, and position in real time. Such information is important for improving traffic safety. In the meantime, it allows vehicles to collaborate with each other, which may significantly improve traffic operations particularly at intersections and freeway ramps. Such collaborations can be facilitated by autonomous vehicles as they require only a minimum “reaction time”. Additionally, such information can be shared with upstream vehicles in a timely fashion. In this way, they can take proactive actions such as moving to the median or to the left-most lane to make room for the merging ramp vehicles. Intuitively, these collaborative driving behaviors may contribute to smoother merging maneuvers and improved operations in freeway ramp areas.

Recently, many studies (1–4) have been conducted to investigate the potential benefits that connected vehicle technologies may bring. Most of these studies are focused on intelligent vehicle and traffic control, traffic safety, advanced traveler information systems, and incident

management. As an important aspect of intelligent traffic control, ramp control has also been addressed in several of these studies that are summarized below.

Sivaraman and Trivedi (5) investigated active traffic safety based on predictive driver assistance (PDA) and cooperation among vehicles, drivers, and infrastructure. Four levels of cooperation strategies were considered, which were in-vehicle cooperation, vehicle-driver cooperation, vehicles to vehicles (V2V) cooperation, and vehicles to infrastructure (V2I) cooperation. The authors used Markov Chain Monte Carlo (MCMC) simulation to quantify the safety effects of PDA and cooperation at various levels based on a predefined near-collision scenario. Four cases consisting of one ramp vehicle merging onto a highway were simulated and analyzed. The results show that active safety was greatly enhanced by either one of the PDA, V2V, or V2I strategies.

Shingde et al. (6) proposed and implemented two algorithms called Head of the Lane (HoL) and All Feasible Sequences (AFS) for automated merge control. The HoL is a distributed and iterative merge control algorithm. In each iteration, two vehicles closest to the merging point from the two merging approaches follow certain rules to determine their merge sequence. The AFS is a centralized algorithm. Instead of using an iterative approach, it takes a snapshot of all vehicles from the two merging approaches and determines their merge sequence simultaneously. Experimental results suggest that HoL and AFS perform equally well in terms of average driving time to intersection (DTTI). HoL works only when the traffic volume is low, while AFS works for both low and high traffic conditions.

Milanés et al. (7) developed a ramp control system consisting of a control model and a fuzzy controller. The control model determines when a merging vehicle should enter the main road. The fuzzy controller is used to “drive” all vehicles following the decisions made by the control model. This system was first tested using a simulator. It was further validated in the real world using three vehicles. The real-world validation considered a congested scenario with two closely spaced vehicles on the main road and the third vehicle on the minor road. Lu and Hedrick (8) also developed a mathematical approach for modeling one vehicle from the minor road merging onto the main road between two vehicles. Although these developed methods worked for the simple scenario with only three vehicles, it is unclear whether they can effectively and safely handle more complicated scenarios with multiple merging vehicles.

Cao et al. (9) proposed a nonlinear model predictive control (MPC) method for merging traffic control. Only two vehicles were considered in their model and case studies, with one on the main road and the other one on the minor road. This merge control problem was formulated as a nonlinear optimization problem and solved by C/GMRES. Similar to previous studies on merge control, this study by Cao et al. considered oversimplified cases and the developed model may not be generalized to solve real-world merge control problems.

Although these studies (5–9) suggest that connected or autonomous vehicles can improve traffic safety and increase traffic throughput at freeway ramps, none of them looked at how to optimally coordinate the movements of freeway and ramp vehicles in a complex and realistic setting. In this paper, a nonlinear optimization model is developed for this purpose. This model takes the second-by-second accelerations of all vehicles as the decision variables and tries to maximize the total speed of all vehicles over the next short time period. It also ensures that when a vehicle arrives at the merging point, the distance headways between it and adjacent vehicles are greater than a minimum value to guarantee safety. In addition, a simulation platform is developed based on VISSIM, MATLAB, and the VISSIM Car2X module to quantify the benefits of this optimal ramp control strategy. The optimization model is detailed in Section 2. Section 3

describes the simulation platform. The proposed model and simulation platform are validated in Section 4 using several case studies. Section 4 also describes the plan to evaluate the proposed optimal control strategy using simulation. The simulation results are presented and discussed in Section 5. Section 6 provides conclusions and recommendations for future research.

MODEL FORMULATION

Figure 1 shows a typical freeway on-ramp, based upon which the proposed control model is formulated. The freeway segment under investigation is about 1,000 m long. The location where the ramp and the freeway connect is called the merging point. A merging zone is defined as a segment of the rightmost lane of the freeway that is within 250 m downstream of the merging point. Under the connected vehicle environment, freeway vehicles will be informed of the traffic conditions in the merging zone and on the ramp. Based on such information, freeway vehicles may slow down, accelerate, or shift to the left lane to allow vehicles from the ramp to join the freeway. They may also choose to do nothing. To model such behaviors, four reaction zones, Zones 0 through 3, are considered in this study. As shown in Figure 1, Zones 1~3 are $[0, 250)$, $[250, 500)$, and $[500, \infty)$ meters upstream of the merging point, respectively. Zone 0 is $[0, \infty)$ meters downstream of the merging point.

This study assumes that all lane changes are completed in Zone 3. Upon entering Zone 2, vehicles switch to the autonomous driving mode until they enter Zone 0. In Zones 2 and 1, freeway vehicles are not allowed to change lanes. Zone 2 is for both freeway and ramp vehicles to adjust their longitudinal trajectories based on the optimal control model described below. Following the optimized trajectories, these vehicles can safely pass the merging point without any conflicts. Upon entering Zone 1, all vehicles should travel at a constant speed. Once these vehicles leave Zone 1, human drivers can take over the vehicle control. The proposed optimal ramp control model is based on a strict assumption that all vehicles are connected via dedicated short-range communications (DSRC). Once freeway and ramp vehicles are in Zone 2, they will turn the control over to a central traffic controller and strictly execute the instructions received from the central controller.

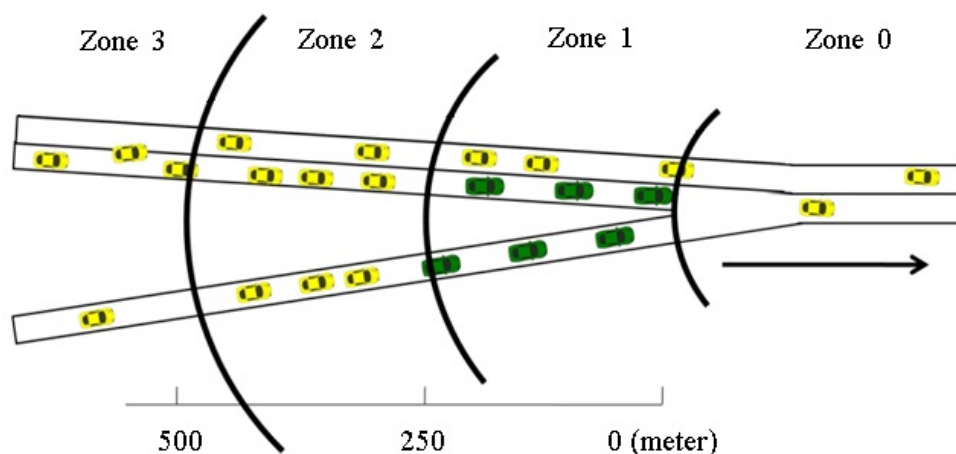


FIGURE 1 Freeway on-ramp considered in this study.

The optimal control strategy is formulated as a nonlinear optimization problem as shown in Equations 1 through 7.

$$\text{Min } \left(-\sum_{i=1}^2 \sum_{j=1}^{n_i} \sum_{k=1}^m v_{i,j,t_k} \right) \quad (1)$$

s.t.

$$0 \leq v_{i,j,t_k} \leq v_{\max} \quad \forall i, j, k \quad (2)$$

$$G_{\min} \leq |x_{i,j,t_k} - x_{i,j-1,t_k}| \quad \forall i, k; j = 2, \dots, n_i \quad (3)$$

$$G_{\min} \leq |x_{1,j,m} - x_{2,p,m}| \quad \forall j = 1, \dots, n_1; p = 1, \dots, n_2 \quad (4)$$

$$|a_{i,j,t_k} - a_{i,j,t_{k+1}}| \leq a_{\max_diff} \quad \forall i, j; k = 1, \dots, m - 1 \quad (5)$$

$$\frac{x_{i,j,t_{k+1}} - x_{i,j,t_k}}{t_{k+1} - t_k} = v_{i,j,t_k}; \quad \frac{v_{i,j,t_{k+1}} - v_{i,j,t_k}}{t_{k+1} - t_k} = a_{i,j,t_k} \quad \forall i, j; k = 1, \dots, m - 1 \quad (6)$$

$$a_{\min} \leq a_{i,j,t_k} \leq a_{\max} \quad \forall i, j; k \quad (7)$$

where

- i = lane identifier (1-ramp and 2-freeway right lane);
- j, p = vehicle index;
- k = time step index;
- m = total number of time steps ($m = 10$ for this study);
- n_i = total number of vehicles in Zone 2 of lane i ;
- t_k = the k^{th} time step;
- a_{i,j,t_k} = acceleration of vehicle j in lane i at time step t_k ;
- v_{i,j,t_k} = velocity of vehicle j in lane i at time step t_k ;
- x_{i,j,t_k} = distance of vehicle j in lane i at time step t_k to the merging point;
- v_{\max} = speed limit;
- G_{\min} = minimum distance gap;
- a_{\min} = minimum acceleration rate;
- a_{\max} = maximum acceleration rate; and
- a_{\max_diff} = maximum acceleration rate change between two consecutive time steps.

A decision interval of 10 s (i.e., $m = 10$) is considered. This interval is further divided into 10 1-s decision steps. At the beginning of each 1-s decision step, each vehicle needs to decide its acceleration rate, which is a decision variable of the above optimization model. By optimizing these acceleration rates, the optimal control model aims to maximize the total speed of all merging vehicles in each decision step subject to the following constraints:

- Constraints 2 ensure that each vehicle maintains a nonnegative speed (v_{i,j,t_k}) that is no greater than the speed limit;

- Constraints 3 require that the distance between two consecutive vehicles in the same lane must be greater than a minimum value G_{min} ;
- Constraints 4 make sure that any pair of freeway and ramp vehicles maintains a safe distance at the end of the decision interval (i.e., when $k = 10$). This is achieved by projecting ramp vehicles onto the freeway using the merging point as the reference;
- Constraints 5 limit the acceleration rate changes of each vehicle between two consecutive time steps to prevent aggressive driving behaviors;
- Constraints 6 describe the relationships among speed, acceleration, and distance. Acceleration is the derivative of velocity with respect to time, and velocity is the derivative of distance traveled to time; and
- Constraints 7 ensure that each vehicle maintains an acceleration rate that is no larger than a_{max} and no less than a_{min} at each time step.

To further prevent aggressive driving behaviors from happening, the original objective function 1 is modified by adding a second term as shown in 8, where $SD_{i,j}$ is the standard deviation of accelerations for vehicle j in lane i . This new term is identical to the notion of acceleration noise which has been used previously in traffic flow control (10). This new objective function has been used to carry out all the case studies and simulations in this research.

$$\min \left(-\sum_{i=1}^2 \sum_{j=1}^{n_i} \sum_{k=1}^m v_{i,j,t_k} + \sum_{i=1}^2 \sum_{j=1}^{n_i} SD_{i,j} \right) \quad (8)$$

DEVELOPMENT OF MODELING FRAMEWORK

To evaluate to what extent the optimal control model can improve traffic operations at freeway on-ramps, an integrated modeling platform is developed. First, a microscopic traffic simulator, VISSIM, is integrated into the platform to simulate the merging process at freeway on-ramps. With the VISSIM simulator and the Car2X module included in it, the accelerations, speeds and positions of all vehicles in Zone 2 can be captured precisely. This information is fed into an optimization module coded in MATLAB. The optimization module takes all the inputs and finds the optimal control strategies (i.e., accelerations) for each vehicle. These optimal strategies are sent back to the VISSIM simulator for vehicle control. To facilitate the data exchange between MATLAB and VISSIM, a C++ application is developed. The optimization module written in MATLAB is encapsulated into a dynamic-link library and called by the C++ application. This C++ application is then compiled as an executable file to override the default driver behavior model in VISSIM. Figure 2 illustrates the architecture of the modeling platform.

MODEL VALIDATION AND EXPERIMENTAL DESIGN

Model Validation

To demonstrate how the optimization-based control algorithm works and to verify the effectiveness of the proposed model, two case studies are conducted. In both studies, the following parameters are used: $v_{max} = 25$ m/s (about 60 mph), $a_{max} = 5$ m/s², $a_{min} = -5$ m/s²,

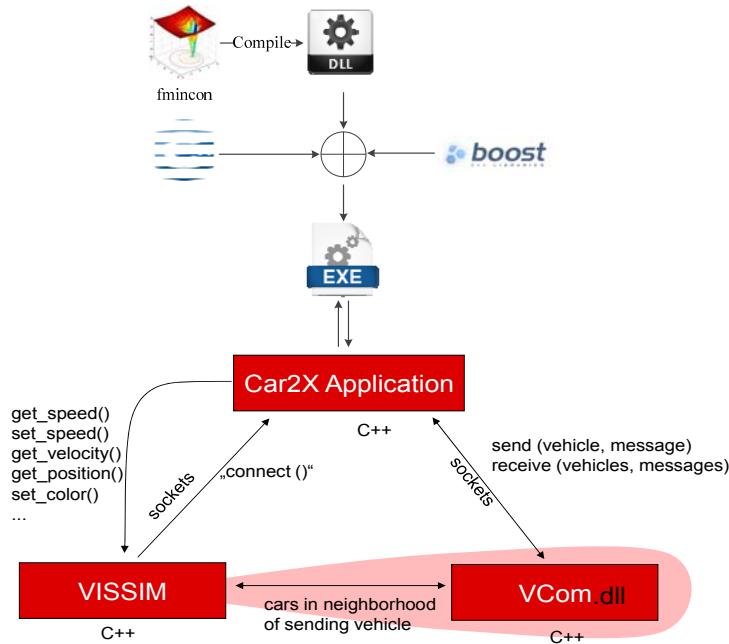


FIGURE 2 Integrated platform architecture.

$a_{\max_diff} = 2 \text{ m/s}^2$, and $G_{\min} = 10 \text{ m}$. Since vehicles are not allowed to change lanes in Zones 2 and 1, only one lane is considered for the freeway in this study.

Case Study I: Four Vehicles

In this case study, two freeway vehicles and two ramp vehicles are considered. The initial accelerations and speeds of all vehicles are assumed to be 0 m/s^2 and 20 m/s , respectively. The initial vehicle states are summarized below. Clearly, if all vehicles maintain their initial speeds, the freeway and ramp vehicles will run into each other at the merging point.

$$\begin{aligned}
 x_{1,j,0} &= [490 \ 500], & x_{2,j,0} &= [490 \ 500], \\
 v_{1,j,0} &= [20 \ 20], & v_{2,j,0} &= [20 \ 20], \\
 a_{1,j,0} &= [0 \ 0], & a_{2,j,0} &= [0 \ 0]
 \end{aligned}$$

The modeling results for this case study are summarized in Figure 3. Veh1-1 and veh1-2 stand for the first and second vehicles on the freeway and veh2-1 and veh2-2 denote the two vehicles on the ramp. Figure 3a clearly shows that the constraints on accelerations are satisfied. Since the initial speeds are all less than the speed limit, the vehicle accelerations are all positive. These vehicles adopt different acceleration patterns in order to maximize their speeds and maintain sufficiently large distance headways at the time they arrive at the merging point. As shown in Figure 3b, none of the four vehicles exceed the 25 m/s speed limit and their speeds all reach 25 m/s at the end of the 10-s decision interval. The time-space diagram in Figure 3c shows the trajectories of the four vehicles. This is done by projecting the four vehicles onto a single lane and using the merging point as the reference point for calculating the distance. At the

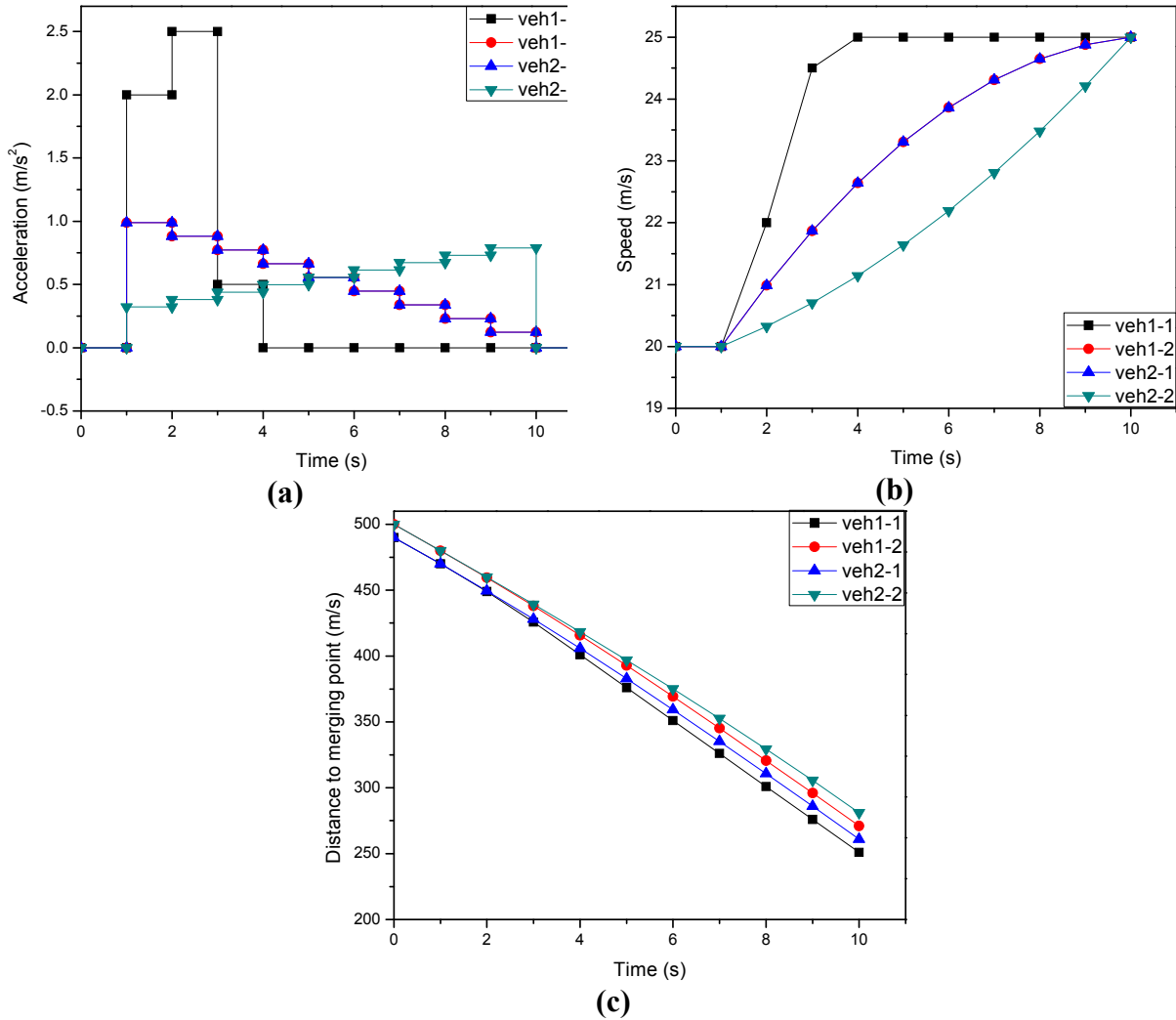


FIGURE 3 Optimization results for Case Study I: (a) acceleration versus time, (b) speed versus time, and (c) distance versus time.

beginning, the distance headway between veh1-1 and veh2-1 is 0. The same thing is true for veh1-2 and veh2-2. By executing the optimal acceleration instructions produced by the model, the four vehicles can pass the merging point safely with a minimum headway of 10 m.

Case Study II: 20 Vehicles

Case study II is used to validate the optimal control model’s performance under heavy traffic conditions. It considers 10 freeway vehicles and 10 vehicles on the ramp in Zone 2. The initial states of all vehicles are

$$\begin{aligned}
 x_{1,j,0} &= [500 \ 490 \ 480 \ 470 \ 460 \ 450 \ 440 \ 430 \ 420 \ 410] \\
 x_{2,j,0} &= [500 \ 490 \ 480 \ 470 \ 460 \ 450 \ 440 \ 430 \ 420 \ 410] \\
 v_{1,j,0} &= [20 \ 20 \ 20 \ 20 \ 20 \ 20 \ 20 \ 20 \ 20 \ 20]
 \end{aligned}$$

$$v_{2,j,0} = [20 \ 20 \ 20 \ 20 \ 20 \ 20 \ 20 \ 20 \ 20 \ 20 \ 20]$$

$$a_{1,j,0} = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

$$a_{2,j,0} = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

The speed and acceleration profiles of the 20 vehicles are presented in Figure 4. Because of the relatively large number of vehicles, the acceleration profiles are mixed together. Compared to those for case study I, the acceleration profiles here are also more difficult to interpret due to the complicated interference among the different vehicles. Nevertheless, the results indicate that constraints 5 and 7 are satisfied. Figures 4c and 4d show that some vehicles may not be able to reach the speed limit (25 m/s) when they pass the merging point due to congestion. These vehicles have to slow down in order to create safe gaps for each other. Figure 5 shows the trajectories of the 20 vehicles. The pattern shown in Figure 5 is very similar to that in Figure 3. If all vehicles are projected onto a virtual lane, at the end of the 10-s interval, one will see a platoon of approximately evenly spaced vehicles.

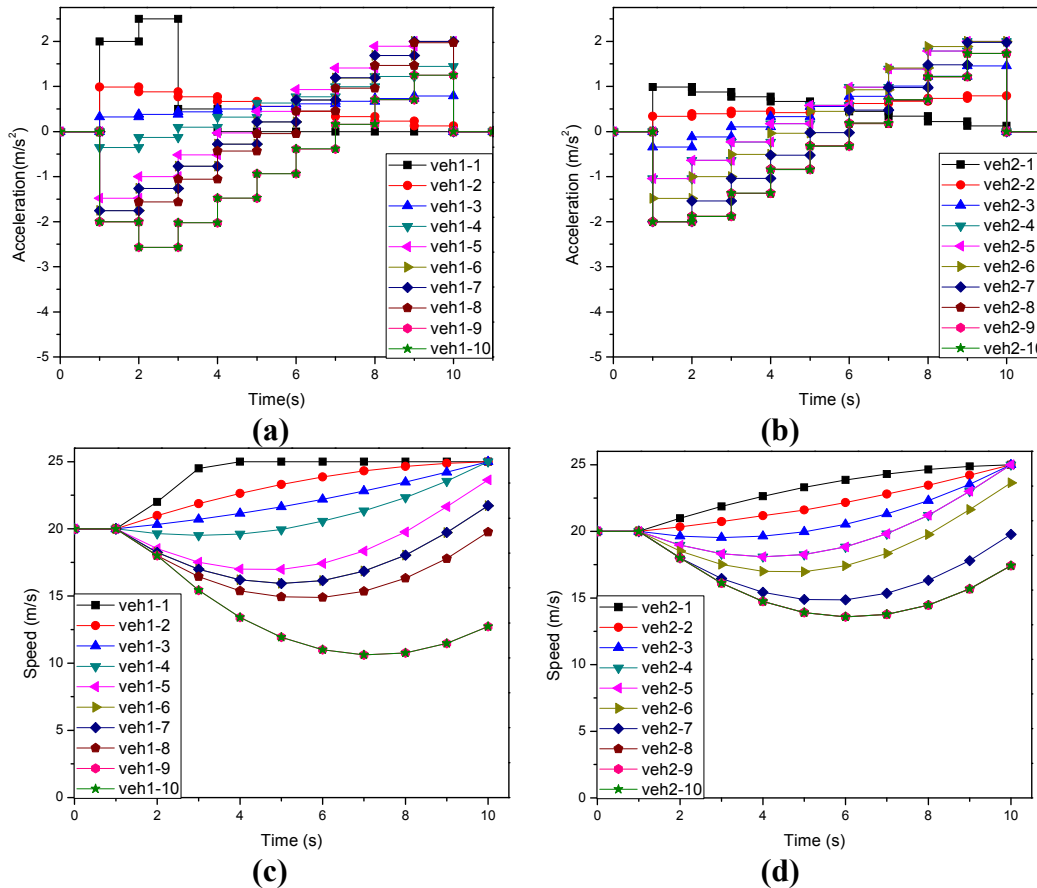


FIGURE 4 Optimized speed and acceleration results for case study II: (a) acceleration versus time for vehicles on highway, (b) acceleration versus time for vehicles on ramp, (c) speed versus time for vehicles on highway, and (d) speed versus time for vehicles on ramp.

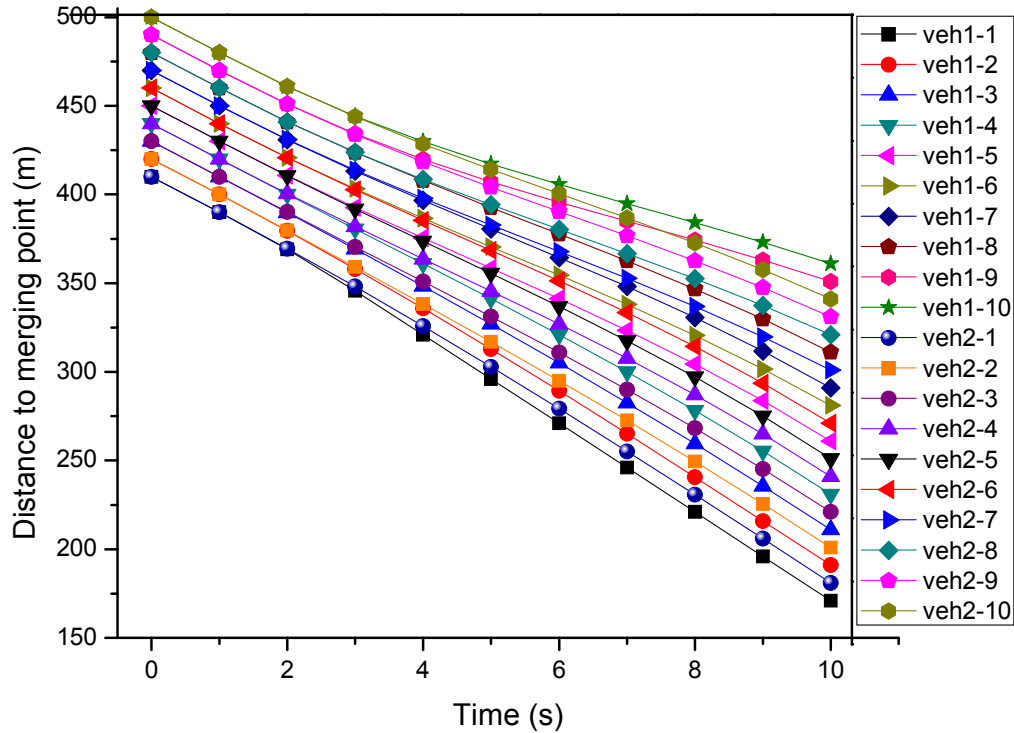


FIGURE 5 Optimized time–distance diagram for Case Study II.

Simulation Framework Validation

This validation study is to examine the integration of VISSIM, MATLAB, and Car2X. It is done by comparing the optimized control instructions generated by MATLAB with the VISSIM control results. For this validation, the freeway and ramp vehicle inputs are set to 1,000 vph and 500 vph, respectively. A random seed of 19 is used in VISSIM to run the simulation. The simulated vehicle trajectories between 30 and 40 s from the beginning of the simulation are recorded and compared with those calculated by the optimal control model coded in MATLAB.

At 30 s, there are seven vehicles in Zone 2. Veh-H-12, Veh-H-13, and Veh-H-14 are on the freeway and Veh-R-15, Veh-R-16, Veh-R-17, and Veh-R-18 are on the ramp. Their distances from the merging point, their speeds and accelerations are plotted in Figure 6. Between 30 and 40 s, these vehicles meet all the speed and acceleration constraints. Table 1 shows the distances of each vehicle to the merging point at different time steps. D1 in Table 1 is the minimum distance between any two vehicles in the same lane. D2 is the minimum distance between any two vehicles from different lanes in Zone 2. The data in Table 1 shows how vehicles adjust their speeds to create safe gaps at the merging point. At 30 s, D2 is 4.09 m, which is between Veh-H-14 and Veh-R-15. Therefore, at least one of these two vehicles should either accelerate or decelerate. At the end of the optimization, D2 has been increased to 10.02 m, which is larger than the required minimum safe distance of 10 m.

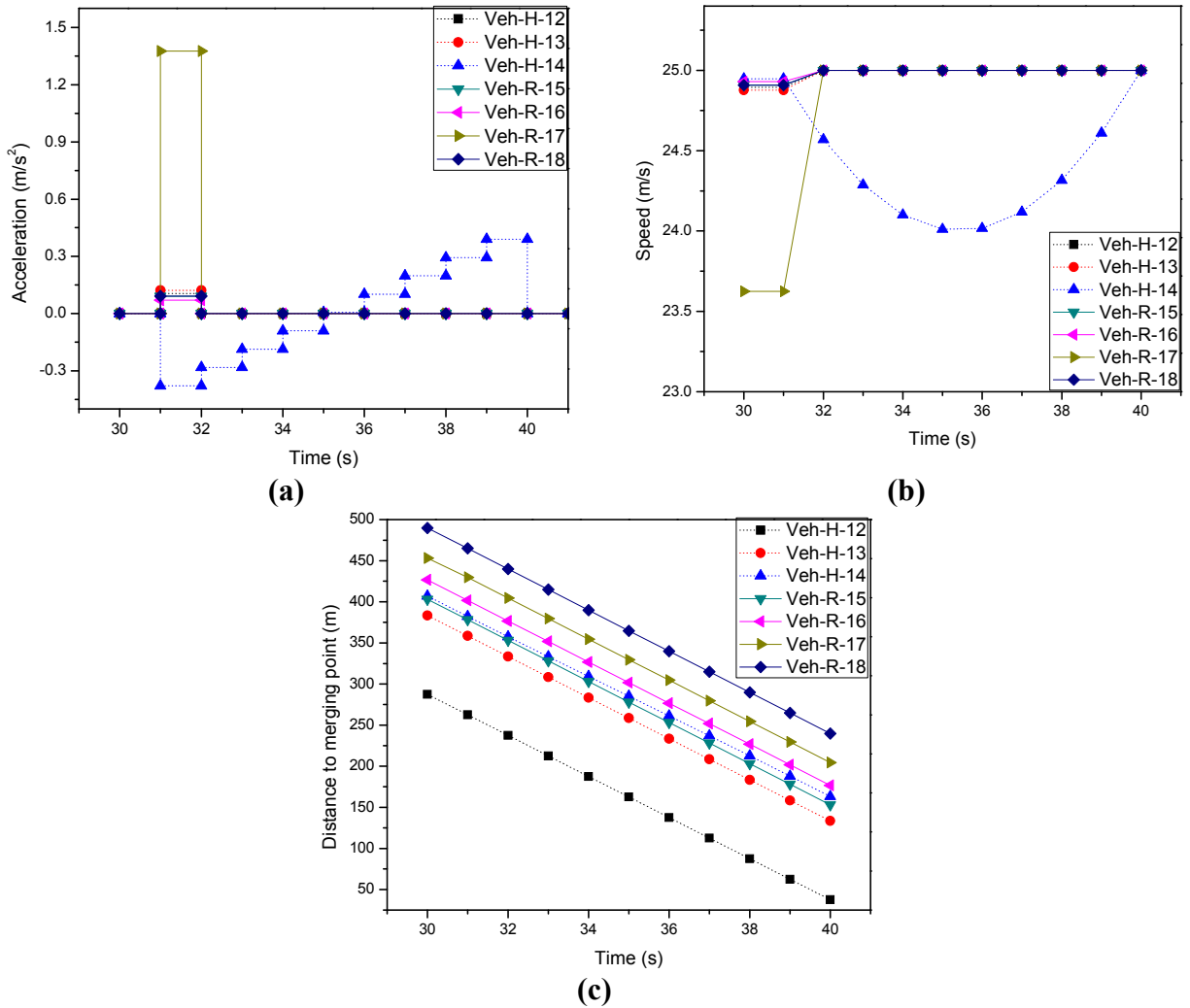


FIGURE 6 Data collected from VISSIM simulation: (a) acceleration versus time. (b) speed versus time, and (c) distance versus time.

TABLE 1 Distance to Merging Point for Vehicles

Time (s)	Distance to Merging Point (m)							D1	D2
	Veh-H-12	Veh-H-13	Veh-H-14	Veh-R-15	Veh-R-16	Veh-R-17	Veh-R-18		
30	287.46	383.36	407.10	403.01	426.66	453.20	489.77	23.75	4.09
31	262.57	358.48	382.15	378.10	401.73	429.57	464.86	23.68	4.05
32	237.57	333.48	357.58	353.10	376.73	404.57	439.86	24.11	4.48
33	212.57	308.48	333.30	328.10	351.73	379.57	414.86	24.82	5.20
34	187.57	283.48	309.19	303.10	326.73	354.57	389.86	25.72	6.09
35	162.57	258.48	285.18	278.10	301.73	329.57	364.86	26.70	7.08
36	137.57	233.48	261.17	253.10	276.73	304.57	339.86	27.69	8.07
37	112.57	208.48	237.05	228.10	251.73	279.57	314.86	28.57	8.95
38	87.57	183.48	212.73	203.10	226.73	254.57	289.86	29.25	9.63
39	62.57	158.48	188.12	178.10	201.73	229.57	264.86	29.64	10.02
40	37.57	133.48	163.12	153.10	176.73	204.57	239.86	29.64	10.02

Table 2 shows the accelerations recorded during the VISSIM simulation. The acceleration value in each time step was recorded at the end of the step (i.e., the acceleration of 0.1054 for Veh-H-12 at 32 s means the acceleration from 31 to 32 s is 0.1054). **Table 3** shows the accelerations suggested by the optimal control model coded in MATLAB. The suggested acceleration is given at the beginning of each time step (i.e., the acceleration of 0.105387 for Veh-H-12 at 31 s is for the time step between 31 and 32 s). Comparing the values in Tables 2 and 3 indicates that VISSIM strictly follows the accelerations obtained by the optimal control model. The maximum difference between the optimized and actually executed accelerations is 0.0001 m/s^2 at 32 s for Veh-H-14, which is negligible.

Experimental Design

The previous validation studies show that the proposed optimal control model and integrated simulation platform work under both low and heavy traffic conditions. To further demonstrate their effectiveness, in the rest of this paper, this optimal control model and platform are applied

TABLE 2 Recorded Accelerations During VISSIM Simulation

Time (s)	Highway (m/s^2)			Ramp (m/s^2)			
	Veh-H-12	Veh-H-13	Veh-H-14	Veh-R-15	Veh-R-16	Veh-R-17	Veh-R-18
30							
31	0	0	0	0	0	0	0
32	0.1054	0.1223	-0.3782	0.0914	0.0701	1.3757	0.0917
33	0	0	-0.2823	0	0	0	0
34	0	0	-0.1862	0	0	0	0
35	0	0	-0.0903	0	0	0	0
36	0	0	0.0057	0	0	0	0
37	0	0	0.1018	0	0	0	0
38	0	0	0.1977	0	0	0	0
39	0	0	0.2937	0	0	0	0
40	0	0	0.3897	0	0	0	0

TABLE 3 Suggested Accelerations Calculated in MATLAB

Time (s)	Highway (m/s^2)			Ramp (m/s^2)			
	Veh-H-12	Veh-H-13	Veh-H-14	Veh-R-15	Veh-R-16	Veh-R-17	Veh-R-18
30							
31	0.105387	0.122316	-0.37824	0.091373	0.070093	1.375658	0.091721
32	-4.4E-35	0	-0.28224	-1.6E-18	0	-2.9E-17	0
33	0	0	-0.18626	-6E-18	-9.9E-32	4.23E-18	2.33E-18
34	0	0	-0.09026	-2.8E-18	9.86E-32	8.05E-17	1.15E-19
35	-4.1E-32	-4.1E-32	0.00572	-1E-17	-8.4E-18	1.04E-16	5.88E-18
36	9.86E-32	0	0.101721	-8.5E-18	-7.9E-18	-8.3E-17	4.88E-18
37	0	0	0.197722	8.1E-19	8.52E-18	3.02E-17	-8.2E-33
38	-3.9E-31	0	0.293723	-6.8E-17	1.02E-18	-1.2E-17	-5.5E-17
39	4.93E-32	-3.9E-31	0.3897	6.21E-17	-7.3E-18	6.32E-17	1.09E-16
40	0	0	0	1.37E-35	0	0	0

to various traffic scenarios against two other control cases. The different control cases to be evaluated and compared are summarized below.

Case 0. As its name suggests, this case does not consider any communications among vehicles or any autonomous vehicles. It simply lets human drivers coordinate their merging behaviors and is introduced as a benchmark (i.e., do nothing). All vehicles follow the default driver behavior models included in VISSIM.

Case 1. If the speed of any vehicles in Zone 0 is less than 45 km/h, upstream vehicles in Zones 1 and 2 will be advised to reduce their speeds to 50km/h and 70km/h, respectively. Since the same distance headway will be considered safer under a lower traffic speed, this strategy may create additional safe gaps and allow more on-ramp traffic to merge onto the freeway.

Case 2. This is the same as the proposed optimal control strategy. Trajectories of vehicles in Zone 2 will be collected every 10 seconds. The information is sent to MATLAB for optimization and the optimized accelerations for each vehicle in the next decision interval (10s long) will be sent back to the related vehicles. These vehicles will then strictly follow these optimized acceleration instructions.

For the comparison, three levels of traffic demand (low, medium, high) are considered for the freeway, which are 800 vph, 1,000 vph, and 1,200 vph. Similarly, the on-ramp traffic flow rate is assumed to be 300 vph, 500 vph, and 700 vph, respectively. In total, there are nine different combinations of traffic demand. In all the simulations, the default Wiedemann 99 car-following model and free lane selection behavior in VISSIM are adopted. In addition, vehicles follow the same constraints as described at the beginning of Section 4.1 except for $G_{min} = 15$ m. Each simulation run represents 3,600 s of ramp operations in the real world.

RESULT ANALYSIS AND CONCLUSION

The three cases are compared in terms of average delay time, average speed, and traffic throughput. In the following sections, the outputs for each of the three measurements of effectiveness are presented and discussed in detail.

Average Delay Time

The average delay time results, grouped by the on-ramp traffic flow, are shown in [Figure 7](#). [Figure 7a](#) shows the average delay time results with the on-ramp traffic flow being 300 vph. [Figures 7b](#) and [7c](#) present the results for the on-ramp traffic flow being 500 vph and 700 vph, respectively.

As shown in [Figure 7a](#), when both the freeway and on-ramp flows are low, there is no significant difference between Cases 0 and 1. As the freeway traffic flow increases, Case 0 significantly outperforms Case 1. This suggests that reducing freeway traffic speed is unnecessary for light ramp traffic. The results in [Figure 7b](#) (ramp traffic = 500 vph) indicates that Case 1 performs significantly better than Case 0 for almost all scenarios except when the freeway traffic flow is 1,200 vph. This is probably because when the freeway traffic is low to medium, it is possible to create additional safe gaps for merging ramp vehicles by reducing the

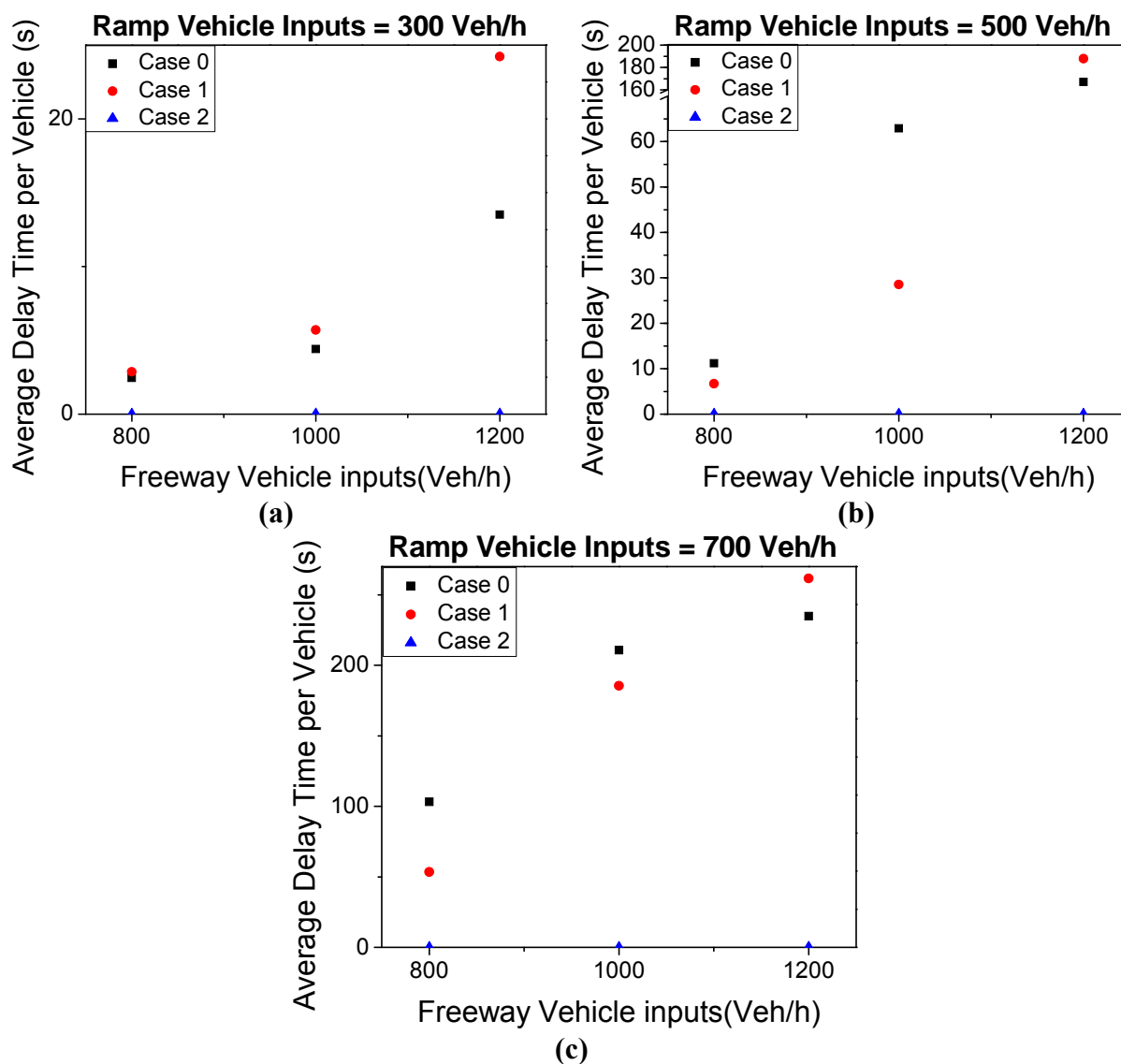


FIGURE 7 Average delay time comparison: (a) low on-ramp traffic, (b) medium on-ramp traffic, and (c) high on-ramp traffic.

speeds of freeway vehicles. However, it is impossible to do so when the freeway traffic is heavy. The results in Figure 7c show the same trend as in Figure 7b. For all traffic flow scenarios considered in this study, Case 2 performs the best and its delay time is almost negligible.

Average Speed

Figure 8 shows the average speed results grouped by the on-ramp traffic flow. For all traffic scenarios considered, Case 2 performs very well and its average speeds are barely affected by the varying traffic flows. On the contrary, the average speeds of both Cases 0 and 1 are significantly reduced by the heavy traffic flows of both ramp and freeway. For high ramp and low freeway flows, Case 1 significantly outperforms Case 0. In this case, reducing the speeds of freeway

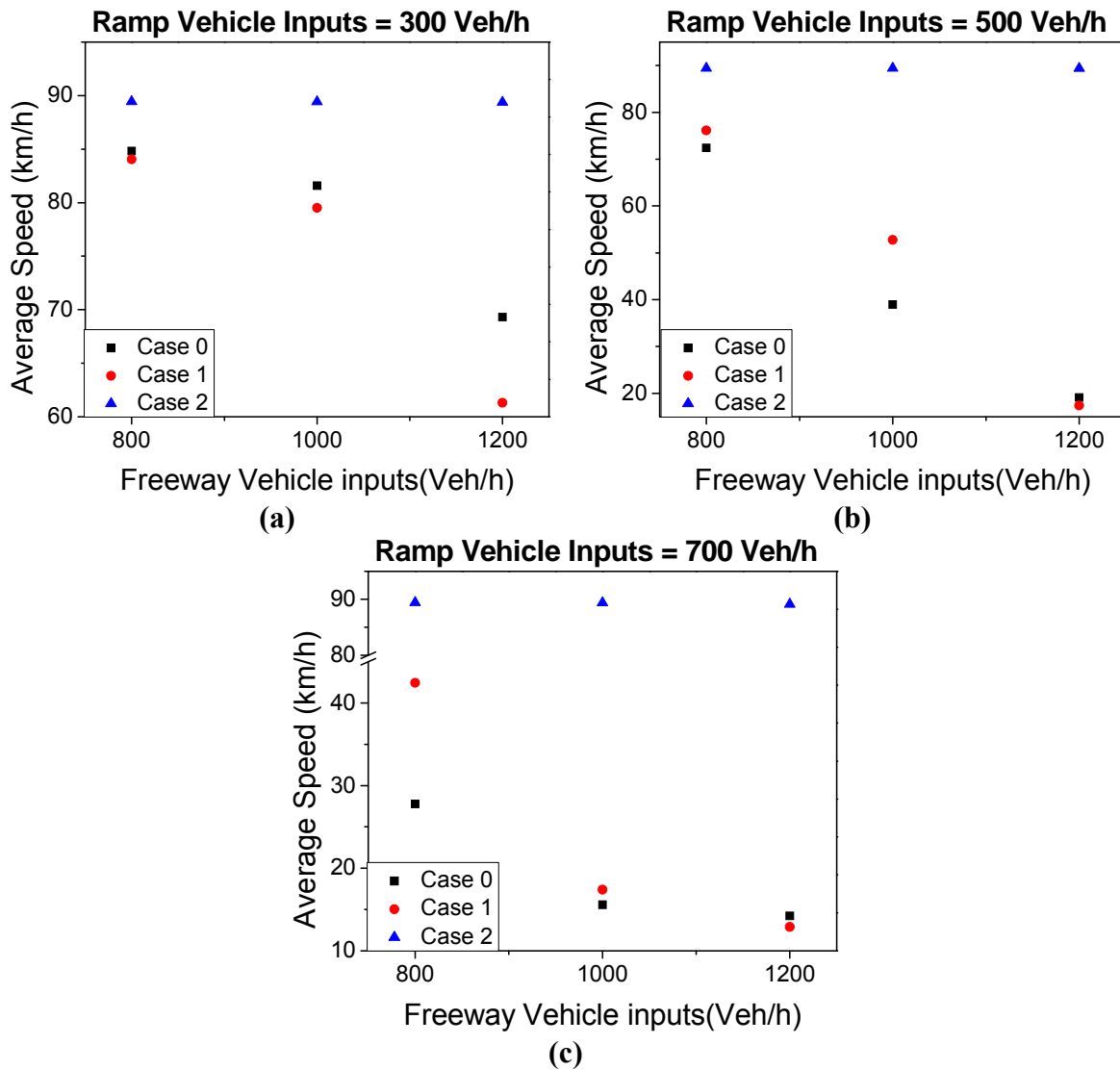


FIGURE 8 Average speed comparison: (a) low on-ramp traffic, (b) medium on-ramp traffic, and (c) high on-ramp traffic.

vehicles can help to create additional safe gaps for ramp vehicles to merge onto the freeway. The overall network average speed thus is increased. However, for low ramp (300 vph) or heavy freeway (1,200 vph) traffic flows, Case 0 consistently performs better than Case 1. This suggests that the gradual speed limit strategy should take into consideration both ramp and freeway traffic conditions.

Traffic Throughput

The traffic throughput results in Figure 9 are consistent with the average delay time and average speed results. When the ramp traffic flow is low (300 vph), there is no major difference among the three cases. As the ramp or freeway traffic flow increases, the differences among the three cases become more significant. In general, Case 2 performs the best and it allows all vehicles to

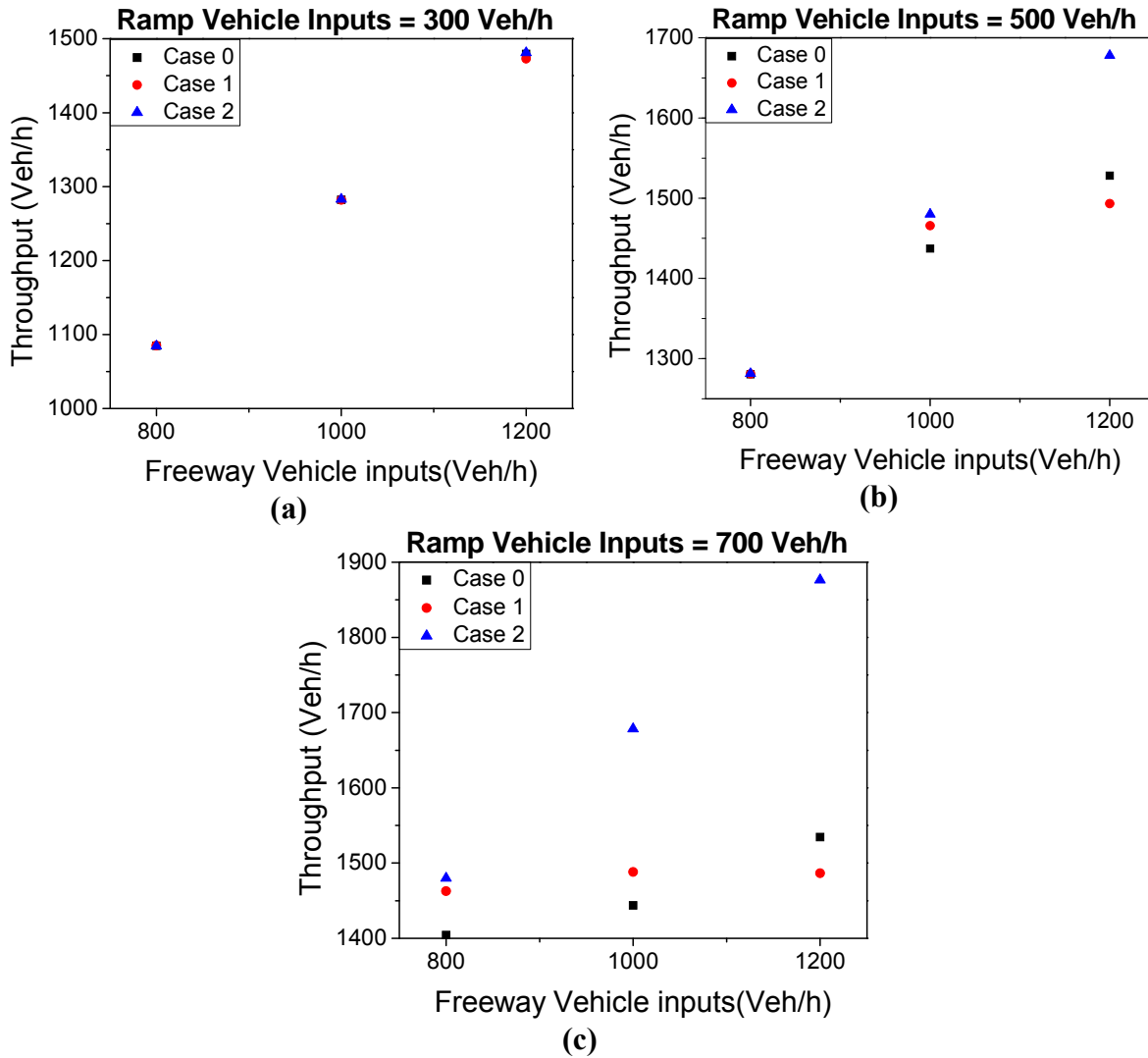


FIGURE 9 Throughput comparison: (a) low on-ramp traffic, (b) medium on-ramp traffic, and (c) high on-ramp traffic.

clear the network for all scenarios. Case 1 performs better than Case 0 when the ramp traffic is heavy and the freeway traffic flow is low. Again, Case 0 outperforms Case 1 when the freeway traffic flow is high and the ramp flow is low.

CONCLUSION

This paper proposes and evaluates an optimization-based ramp control strategy assuming all vehicles are connected and controlled automatically. A simulation platform is developed integrating VISSIM, MATLAB, and the Car2X module in VISSIM. The proposed optimal ramp control strategy is formulated as a nonlinear optimization problem and solved using the MATLAB optimization toolbox. This optimization model divides the decision interval into 1-s time steps. Based on the initial speeds, accelerations, and locations of all vehicles, the control

algorithm takes the second-by-second accelerations of each vehicle as the decision variable and optimizes them. The optimized accelerations are then used to control these vehicles during the next decision interval.

Three case studies are conducted to validate the effectiveness of the developed optimal control model and the simulation platform. The proposed optimal control algorithm (Case 2) is further compared with a do-nothing strategy (Case 0) and a gradual speed limit strategy (Case 1) for controlling a typical freeway on-ramp. Various levels of freeway and on-ramp traffic flows are considered, which results in nine test scenarios. These three ramp control strategies are compared in terms of average delay time, average speed, and traffic throughput. When either the freeway or the on-ramp traffic flow is low, there is no significant difference among the three control strategies in terms of throughput. This is likely because ramp vehicles can all find a safe gap to join the freeway without causing long standing queues. For the remaining scenarios considered, the optimal control strategy substantially outperforms the other two strategies. When the freeway traffic is heavy and the on-ramp traffic is light, the gradual speed limit strategy performs even worse than not considering any control. This gradual speed limit strategy works when the freeway traffic flow is low and the on-ramp has a medium to heavy traffic.

The results demonstrate the potential effectiveness of the proposed optimization-based ramp control strategy. However, it is based on a strict assumption that all vehicles are connected and controlled automatically. In future studies, it would be interesting to consider how to optimally control a mixture of autonomous vehicles and vehicles controlled by human drivers. Also, multilane freeways can be considered instead of a single-lane freeway. In this case, lane changing decisions can be included as decision variables in addition to the acceleration rates.

ACKNOWLEDGMENT

This research is partially supported by the National Institute of Food and Agriculture, United States Department of Agriculture through a grant with South Carolina State University, under the project entitled Enhancing the Disaster Resilience for Rural Communities Through Better Disaster Preparedness and Improved Emergency Response. The authors also thank the University of Massachusetts Lowell Office of the Vice Provost for Research for providing a portion of the financial support.

REFERENCES

1. Yang, Y., and R. Bagrodia. Simulation-Based Study on a Lane-Based Signal System for Merge Control at Freeway Work Zones. *Proc., Vehicular Ad Hoc Networks*, Beijing, China, 2009.
2. Harb, R., E. Radwan, and V. V. Dixit. Comparing Three Lane Merging Schemes for Short-Term Work Zones: A Simulation Study. ISRN Civil Engineering, 2012.
3. Schumacher, H., C. Priemer, and E. N. Slotke. A Simulation Study of Traffic Efficiency Improvement Based on Car-To-X Communication. Presented at VANET '09, Beijing, China, 2009.
4. Srinivas, S., Y. Xie, T. Chigan, and T. Arsava. Review of Advances in Connected Vehicle Research. Working paper, 2013.
5. Sivaraman, S., and M. M. Trivedi. Towards Cooperative, Predictive Driver Assistance. Presented at 16th International IEEE Annual Conference on Intelligent Transportation Systems, 2013, pp. 1719–1924.

6. Shingde, V., G. Raravi, A. Gudhe, P. Goyal, and K. Ramamritham. Merge-by-Wire: Algorithms and System Support. Presented at 2008 Real-Time Systems Symposium, 2008.
7. Milanes, V., J. Godoy, J. Villagra, and J. Perez. Automated On-Ramp Merging System for Congested Traffic Situations. *IEEE Transactions on Intelligent Transportation Systems*, 2011, pp. 500–508.
8. Lu, X. Y., and J. K. Hedrick. Longitudinal Control Algorithm for Automated Vehicle Merging. *International Journal of Control*, Vol. 76, No. 2, 2003, pp. 193–202.
9. Cao, W., M. Mukai, T. Kawabe, H. Nishira, and N. Fujiki. Automotive Longitudinal Speed Pattern Generation With Acceleration Constraints Aiming at Mild Merging Using Model Predictive Control Method. Presented at 9th Asian Control Conference (ASCC), Istanbul, 2013.
10. Chung, C. C., and N. H. Gartner. Acceleration Noise as a Measure of Effectiveness in the Operation of Traffic Control System. Working paper OR 015-73. Operations Research Center, MIT, March 1973.

POSTER PRESENTATION

A Second-Order Lagrangian Macroscopic Traffic Flow Model for Freeways

ZHUOYANG ZHOU

PITU MIRCHANDANI

Arizona State University

A macroscopic traffic flow model describes the evolution of aggregated traffic characteristics over time and space, which is a basic and critical component for various modern intelligent transportation systems, e.g., a real-time freeway control system. Traditional macroscopic traffic flow models are built in the Eulerian coordinates using Eulerian traffic characteristics, such as density, speed and flow. Recently, the Lagrangian traffic flow modeling using Lagrangian traffic characteristics, such as spacing and speed, began to attract research attentions. It is favored over the Eulerian model mostly for its more accurate and simplified discrete simulation results (1, 2), and its convenience in incorporating vehicle-based information. However, up to now only a first-order model is presented (1, 2). Our paper proposes a new second-order Lagrangian macroscopic traffic flow model for freeways. The idea originates from Payne's second-order Eulerian model which was derived on the basis of car-following considerations (3). It reflects the fact that a driver usually adjusts the speed based on the traffic condition ahead, and that a time delay exists as a driver reacts to the changed traffic condition. A dynamic speed equation is formulated to represent these behavioral facts. A lane-drop scenario is simulated to examine the model performance. Comparison with the first-order Lagrangian model confirms the effectiveness and advantages of the proposed second-order model.

INTRODUCTION

For over 60 years, researchers have devoted great efforts to model the complicated real-world traffic flow system in mathematical ways. Macroscopic traffic flow modeling is a representative way to describe the evolution of aggregated traffic characteristics over time and space. The macroscopic models are widely applied in various modern intelligent transportation systems (ITS). For example, in real-time traffic state estimation, model-based traffic state estimators use macroscopic models to propagate aggregated traffic state variables over time and space.

According to the levels of detail, traffic flow modeling can be categorized into three levels. The lowest aggregated level depicting detailed individual driving behaviors is referred to as the microscopic traffic flow modeling, and the highest aggregated level focusing on the average traffic flow characteristics while neglecting the heterogeneous driving behaviors is referred as the macroscopic traffic flow modeling. Mesoscopic traffic modeling lies in between, digging into more details than the macroscopic models but still leaving out the detailed driving behaviors such as lane changing. Our research focuses on the macroscopic level.

With respect to coordinate systems, macroscopic traffic flow models can fall into either Eulerian models or Lagrangian models. The Eulerian coordinate system has its origin fixed in a spatial position, in which an observer standing on the freeway is stationary in this system while

an observer riding on a vehicle is moving. In a Lagrangian coordinate system, the origin is fixed on the moving vehicles and thus any observer stationary to the freeway (stationary in the Eulerian coordinate system) is moving. The difference of the two systems lies in the change of observing perspectives. Typically, an Eulerian model defines state variables as locally aggregated traffic flow characteristics (traffic density, traffic flow, mean space speed), while a Lagrangian model uses vehicle group based characteristics (vehicle group position, vehicle group mean speed, vehicle group mean spacing). In this paper, we develop our model in the Lagrangian coordinate system, which is a relatively new and promising area.

Macroscopic traffic flow models can also be classified into first-order or second-order (or higher-order) models. The conservation equation and a fundamental diagram constitute a first-order model, while a second-order model incorporates an additional dynamic speed equation. In other words, the speed is no longer statically determined by the density–spacing condition according to the fundamental diagram. With the dynamic speed extension, researchers intend to elaborate the models on their capabilities in reproducing some of the real-world traffic patterns like oscillation in congestions, traffic hysteresis, and gradual speed change. In addition, the second-order model can be conveniently used in joint estimation of both density–spacing and speed at the same time in model-based real-time traffic state estimation, which further facilitates the utilization of speed measurements to enhance the estimation accuracy.

Traffic flow models may also have different emphasis on physical road conditions, e.g., freeways or local routes. Our proposed model assumes a freeway stretch.

There are several reasons motivate us to propose this second-order Lagrangian traffic flow model: (a) for empirical considerations, we believe that the reaction delay and anticipation should be reflected in the speed equation; (b) for practical considerations, this model would be a good candidate to conveniently incorporate vehicle-based spacing and speed information in model-based traffic state estimation; and (c) for theoretical considerations, there are several existing Eulerian second-order models serving as a complementation to the first-order models. However, no such a model exists in the Lagrangian framework. Our model is the result of the efforts in bridging this gap.

For these purposes, a second-order Lagrangian macroscopic traffic flow model for freeways is proposed. In the rest of this paper, we first present an overview of existing traffic flow models in Section 2. Details of the proposed model are illustrated in Section 3. We evaluate the performance of this model in Section 4 through a simulation-based experimental analysis and conclude this paper in Section 5.

LITERATURE REVIEW AND MOTIVATIONS

A macroscopic traffic flow model generally includes (a) a conservation equation and (b) a static or dynamic speed equation. This section reviews some popular existing Eulerian and Lagrangian traffic flow models.

Eulerian Traffic Flow Models

Traditional macroscopic traffic flow models are built in the Eulerian coordinate system. The Lighthill-Whitham and Richards (LWR) model (4, 5) is the most well-known model serving as

the basis for most current macroscopic traffic flow models. It models the traffic flow dynamics as a kinematic wave model described by the partial differential equations below.

$$\begin{cases} \frac{\partial q(x,t)}{\partial x} + \frac{\partial \rho(x,t)}{\partial t} = 0 & (a) \\ q(x,t) = Q[\rho(x,t)] & (b) \end{cases} \quad (1)$$

Equation 1a is the conservation equation, stating that the change in flow (q) over space (x) equals to the change of density (ρ) over time (t) in the $x \times t$ space. Equation 1b is the fundamental diagram describing the nonlinear relationship between density and flow. Almost all of the macroscopic models are developed from this LWR model. For example, the widely applied cell transmission model (CTM) (6, 7) is a discretized version of LWR with the assumption that the fundamental relationship between traffic flow and density is triangular. The freeway segment is discretized into a chain of fixed-length cells, assuming that the traffic condition within each cell is homogeneous. At each time step the traffic state variables are propagated along the spatial “cell-chain.”

With specific formulation of the fundamental diagram and some numerical computation method, the LWR-type models can be implemented in discrete simulations. Plenty of research has been conducted in this area to demonstrate the capabilities and drawbacks of the LWR-type models. It is demonstrated to be capable of reproducing basic phenomena observed in dynamic traffic flows in the real world, such as the conservation of vehicles, the forming and dissolving of congestion at bottlenecks, and the fact that the traffic condition is propagated over time and space in different directions under different prevailing traffic modes (e.g., congested or free flow) (6, 8). However, the main drawback of the LWR-type models is that vehicles are assumed to be able to attain their desired speed (represented by the fundamental diagram) instantaneously, implying infinite acceleration and deceleration. This problem and the resulted simulation inaccuracy lead the researchers to build higher-order models, mostly by incorporating a dynamic speed equation.

The second-order models assume that before vehicles reach the equilibrium speed described by the fundamental diagram, there is a time delay for acceleration or deceleration. For example, Payne (3) formulates the dynamic speed–density relationship as

$$V(x,t + \tau) = V[\rho(x + \Delta x,t)] \quad (2)$$

Equation 3 has the following assumptions: First, drivers adjust their speeds according to the downstream traffic condition, and thus the space mean speed $v(x, t)$ depends on the downstream density $\rho(x + \Delta x,t)$. Second, drivers gradually accelerate/decelerate in reaction to the traffic condition change ahead, that is, the desired (equilibrium) speed would be reached with a time delay (τ). Third, the dynamic speed relationship can be described by the Eulerian fundamental diagram with perturbations τ and Δx . Indeed, this model was derived on the basis of car-following considerations. Papageorgiou (9) extended the model to account for ramp flows. As we mentioned previously, the second-order models have been shown to be capable of reproducing traffic patterns where the first-order models are incapable of, and serving as a good candidate model for real-time traffic state estimations (10). However, there are also arguments questioning the reasonability of the high-order models. Daganzo criticizes the possibility of negative flow and speed prediction in some extreme conditions (11). Others have noticed the

underlying assumption that the desired speed distribution is a property of the road but not the drivers (12). Papageorgiou argues that the second-order models have their merits in remedying the false assumption of the static fundamental diagram, and that the negatively predicted flow–speed values occur only in extreme conditions (13). Other common blames include that the second-order models have more parameters to be calibrated, and that the stability requirement becomes stronger in discrete simulation. Fortunately, these difficulties can be successfully overcome with careful model calibration efforts. As we see more and more second-order models in ITS applications in recently years, we believe that the advantages of the second-order models cannot be dismissed.

Lagrangian Traffic Flow Models

Originating from the flow dynamics theory, macroscopic flow movement can also be observed in the perspective of a moving particle in the flow, which defines the Lagrangian coordinate system. Leclercq (1) presented the formulation of the LWR partial differential equation (PDE) model in Lagrangian coordinates:

$$\begin{cases} \frac{\partial s(n,t)}{\partial t} + \frac{\partial u(n,t)}{\partial n} = 0 & (a) \\ u(n,t) = U[s(n,t)] & (b) \end{cases} \quad (3)$$

The first Equation 2a is the Lagrangian conservation equation. It states that the change of spacing (s) over time equals to the change of the mean speed (u) over the cumulative of number of vehicles (n) in the $n \times t$ space. This Lagrangian conservation equation can be derived from the Eulerian conservation equation by applying $s = 1/\rho$, $q = \partial n / \partial t$, and $\rho = -\partial n / \partial x$ (1). The second Equation 2b is the Lagrangian fundamental diagram, a function relating the spacing to speed, which can be obtained by applying $s = 1/\rho$ to the Eulerian fundamental diagram. An examples of the Lagrangian fundamental diagram is the Smulders' fundamental diagram (14):

$$u = U(s) = \begin{cases} v_f - s_{cr} \cdot \frac{v_f - v_{cr}}{s}, & \text{if } s \geq s_{cr} \\ v_{cr} \cdot \frac{s - s_{jam}}{s_{cr} - s_{jam}}, & \text{otherwise} \end{cases} \quad (4)$$

where v_f , v_{cr} , s_{cr} , and s_{jam} stand for the free-flow speed, the critical speed, the critical density, and the jam density, respectively. Another example is an exponential-form fundamental diagram (9) in Equation 5, where a is a parameter needs calibration.

$$u = U(s) = v_f \cdot \exp \left[-\frac{1}{a} \left(\frac{s_{cr}}{s} \right)^a \right] \quad (5)$$

Van Wageningen-Kessels (2) develops a concrete framework for implementing the first-order Lagrangian traffic flow model for mixed-class and multiclass vehicles. In the mixed-class model, the traffic flow is discretized into vehicle groups using Godunov scheme. The advantages of this alternative to the traditional Eulerian models can be summarized as following. First, the traffic characteristics are propagated only in the flow direction since drivers only react to the front vehicles, which simplifies the discretization procedure (the upwind method). Second, the discrete simulation results show higher accuracy in many scenarios, compared with the Eulerian case. Third, the vehicle-based information (e.g., spacing measurements) can be directly used in state estimation (15). Since the Lagrangian modeling is a relatively new area, no second-order traffic flow model has been presented yet.

A SECOND-ORDER LAGRANGIAN MODEL

Applying the same idea as Payne's (3), we start our modeling with the following dynamic speed equation:

$$u(n,t + \delta_t) = U[s(n - \delta_n, t)] \quad (6)$$

Equation 6 has the following three assumptions. First, the speed is adjusted according to the traffic condition ahead. As n stands for the accumulated volume of vehicles, it is decreasing along the flow direction, and $(n - \delta_n)$ corresponds to a position in the downstream ($\delta_n > 0$). Second, the speed is gradually changing with finite acceleration/deceleration, and the equilibrium speed cannot be reached until δ_t time later. Third, the dynamic speed relationship can be described by the spacing-speed fundamental diagram with perturbations δ_n and δ_t .

Derivation of the Dynamic Speed PDE

Expanding the right side of Equation 6 using Taylor expansion, we obtain Equation 7.

$$\begin{aligned} u(n,t) + \delta_t \cdot \frac{du(n,t)}{dt} &= U[s(n,t)] - \delta_n \cdot \frac{dU[s(n,t)]}{dn} \\ &= U[s(n,t)] - \delta_n \cdot \frac{\partial U[s(n,t)]}{\partial s(n,t)} \cdot \frac{\partial s(n,t)}{\partial n} \end{aligned} \quad (7)$$

We apply the following assumptions that are similar to those in the Eulerian models. Assume $\delta_n = \theta \cdot \rho = \theta / s$, where θ is a constant parameter assumed to be around $\theta = 0.5$.

Assume that $\partial U[s(n,t)] / \partial s(n,t)$ is positively constant. Let $\lambda = \theta \cdot \frac{\partial U[s(n,t)]}{\partial s(n,t)}$, and the right-

hand side of the Equation 7 becomes $U[s(n,t)] - \frac{\lambda}{s} \cdot \frac{\partial s(n,t)}{\partial n}$.

The term $\frac{du(n,t)}{dt}$ in Equation 7 is the acceleration term which should be treated as $\frac{du(n(t),t)}{dt}$, leading to Equation 8.

$$\frac{du(n(t),t)}{dt} = \frac{\partial u}{\partial t} + \frac{\partial u}{\partial n} \cdot \frac{\partial n}{\partial t} \quad (8)$$

Applying $\partial n / \partial t = q = u / s$ and letting $\tau = \delta_t$, we finally obtain the dynamic speed PDE in Equation 9.

$$\frac{\partial u(n,t)}{\partial t} = -\frac{\partial u(n,t)}{\partial n} \cdot \frac{u(n,t)}{s(n,t)} + \frac{1}{\tau} \cdot \left[U[s(n,t)] - u(n,t) - \frac{\lambda}{s} \cdot \frac{\partial s(n,t)}{\partial n} \right] \quad (9)$$

Equation 3a, Equation 9, and a fundamental diagram (e.g., Equation 4, Equation 5) constitute the proposed continuous second-order Lagrangian traffic flow model.

Discrete Model

Let j be the index for the vehicle group and $j = 1$ represents the vehicle group in the most downstream, and k be the index for the time steps. Let Δn denote the vehicle group size over all lanes, and let η_j denote the vehicle group size per lane. Thus $\eta_j = \frac{\Delta n}{L_j}$ and $\partial n = -\eta_j$. Note that in

the macroscopic level n is a continuous dimension, and thus Δn is continuous as well. ε is added to the denominator to ensure computation stability. The discretized model consists of the following three equations.

$$\left\{ \begin{array}{l} s_j(k+1) = s_j(k) + \frac{T}{\eta_j} \cdot [u_{j-1}(k) - u_j(k)] \quad (10) \\ u_j(k+1) = u_j(k) + \frac{T}{\tau} \cdot [U(s_j(k)) - u_j(k)] + \frac{T}{\eta_j} \cdot \frac{u_j(k) [u_{j-1}(k) - u_j(k)]}{s_j(k) + \varepsilon} + \frac{T\lambda}{\tau\eta_j} \cdot \frac{[s_{j-1}(k) - s_j(k)]}{s_j(k) + \varepsilon} \quad (11) \\ U(s_j(k)) = \begin{cases} v_f - s_{cr} \cdot \frac{v_f - v_{cr}}{s_j(k)}, & \text{if } s \geq s_{cr} \\ v_{cr} \cdot \frac{s_j(k) - s_{jam}}{s_{cr} - s_{jam}}, & \text{otherwise} \end{cases} \quad (12) \end{array} \right.$$

Equation 10 is the discrete Lagrangian conservation equation, Equation 11 is the discrete Lagrangian dynamic speed equation and Equation 12 is the Lagrangian Smulders' fundamental diagram, which could also be other valid forms.

The proposed discrete dynamic speed equation (Equation 11) has four components on the right-hand side. The first component is the current speed, the second component represents the influence of the difference between the desired speed and the current speed, the third component represents the influence of the speed of the front vehicle group, and the last component represents the influence of the spacing of the front vehicle group. The last three components together state that the drivers' desire to reach the equilibrium speed and the influences from the leading vehicles result in the finite acceleration/deceleration (i.e., the gradual change of speed) from the current speed. In contrast, the first-order model determines the speed purely by the spacing of the current vehicle group during the same time step. Obviously, the dynamic speed equation is more realistic according to the real-world observations.

The model has three additional critical parameters, i.e., τ , $\lambda(\theta)$, and ε . ε is added to ensure the computation stability, while τ and $\lambda(\theta)$ have physical meanings. τ represents the reaction delay time, and θ represents how much the current vehicle group is influenced by the front spacing condition. Thus, with a larger τ and a smaller θ , one may expect less sensitivity of the drivers to the front condition and more oscillations in congestions. This expectation is observed in the simulation results in Section 4.

This discrete model can be viewed as a state space model, and the state variables are the spacing and speed of the vehicle groups. This formulation makes it convenient to conduct joint estimation of speed and to incorporate vehicle speeds as measurements.

Influence of Merges and Diverges

The model described by Equations 10 through 12 is a freeway link model, i.e., it does not take into considerations of the ramp flows. Van Wageningen-Kessels (2) has shown that with merge (onramp) flows and diverge (offramp) flows, the Lagrangian conservation equation becomes:

$$\frac{\partial s}{\partial t} - \frac{\partial u}{\partial n} = -s^2 \cdot (\alpha - \beta). \quad (13)$$

where α and β stands for the inflow rate and outflow rate [in veh/(m/s)], respectively. The ramp flows can indirectly influence the mainstream speed through this conservation equation. However, the dynamic speed equation also needs to be modified to directly reflect the influence of ramp flows. Generally, the onramp flow merges into the mainstream flow with a lower speed $v_{\text{onramp}} < v_{\text{main}}$ and need to accelerate to match the mainstream speed. For the offramps, the exiting flow needs to decelerate to match the offramp speed limit. Thus both merge and diverge cause speed reduction in the mainstream flow. Previously, an additional term to the dynamic speed equation has been proposed (9) in the Eulerian second-order models to reflect this acceleration/deceleration phenomenon. For onramps it is described as $(v_{\text{onramp}} - v) \alpha / \rho$. Assume $v_{\text{onramp}} = (1 - v) v$, it becomes $-v \cdot v \cdot \alpha / \rho$, $v \in [0, 1]$. For off-ramps it is a similar term with α replaced by β . For the Lagrangian dynamic speed equation, a similar term can be added to Equation 9 with ρ replaced by s , described in Equation 14.

$$\frac{\partial u}{\partial t} = -\frac{\partial u}{\partial n} \cdot \frac{u}{s} + \frac{1}{\tau} \left[U(s) - u - \frac{\lambda}{s} \cdot \frac{\partial s}{\partial n} \right] - (v_\alpha \cdot \alpha + v_\beta \cdot \beta) \cdot v \cdot s \quad (14)$$

The additional term represents the acceleration–deceleration phenomenon, which is in unit of m/s^2 .

Boundary Conditions

The model is not completed for discrete simulation without appropriate boundary conditions. For the inflow boundary the minimum supply–demand method is applied (Equation 15) (16). Let D_i stand for the demand of traffic flow in freeway segment i , and S_i the supply. Let q_{cr} stand for the critical flow. The time step index k is omitted for simplicity. The minimum supply–demand method for the inflow boundary is the same as that for the Eulerian case except for that the freeway segment i is in fact the available space in the most upstream, i.e., the distance between the starting position of the freeway and the position of the last vehicle group that is already in the computation domain. Segment $i - 1$ is a dummy segment where the unsatisfied traffic demand queues. Once the calculated flow $q_{(i-1 \rightarrow i)}$ is as large as a vehicle group of size Δn , it is allowed to enter the computation domain.

$$\begin{aligned}
 q_{(i-1 \rightarrow i)} &= \min D_{i-1}, S_i \\
 D_i &= \begin{cases} q_i & \text{if free flow, } s_i > s_{cr} \\ q_{cr} & \text{o.w.} \end{cases} \\
 S_i &= \begin{cases} q_{cr} & \text{if free flow, } s_i > s_{cr} \\ q_i & \text{o.w.} \end{cases}
 \end{aligned} \tag{15}$$

At the outflow boundary, a dummy vehicle group is assumed to represent the out-boundary condition. The first vehicle group within the computation domain follows the dummy vehicle group according to the upwind method until it leaves the computation domain and becomes the new dummy vehicle group.

Stability Requirements

For the discrete simulation of the first-order Lagrangian model, the Courant-Friedrichs-Lewy (CFL) condition described in Equation 16 ensures the numerical computation stability.

$$\frac{T}{\eta_{\min}} \cdot \max \left| \frac{\partial U(s_j(k))}{\partial s} \right| \leq 1 \tag{16}$$

Indeed, it limits the maximum distance that a vehicle group can travel within one time step. For detailed explanation we refer to (2). The minimum size of vehicle groups per lane (η_{\min}) is obtained by letting $\text{CFL} = 1$. However, the second-order model incorporates a dynamic speed equation, and thus the CFL number cannot be approximated using Equation 16. The condition is in fact the same that during one time-step, a vehicle group cannot cross two discretization boundaries in the numerical computation. Experimental experience from simulation found that the required vehicle group size is larger but still reasonable for a macroscopic level simulation.

Discussions

Arguments between the first-order models and the second-order models have never stopped. As mathematical models are approximations of the physical world, they can hardly be exactly accurate. The same is true for the Lagrangian traffic flow models. There is no simple conclusion for the comparison, as they both have their own advantages and disadvantages. The fundamental diagram is only a coarse approximation of the empirical evidence. The proposed model is more sensible as it improves the first-order model by dynamically modeling the speed. Particularly, the speed property would be related to vehicle groups instead of a location on the road. This makes the Lagrangian second-order model more reasonable than the Eulerian counterpart.

EXPERIMENTAL ANALYSIS

This section presents the experimental analysis of the proposed second-order model using a lane-drop scenario shown in Figure 1. This experimental analysis is not intended to be a quantitative measure of how accurate the model is compared with a real data set. Rather, it is compared with the outputs of the first-order Lagrangian model to show the performance and the potential advantages of the second-order model, with the underlying assumption that the first-order model is reasonable (2).

Scenario Description

A lane-drop scenario (Figure 1) is assumed to emulate the onset and discharge of congestion. The total length of the freeway stretch is 6km without ramps. At $x = 3$ km the number of lanes drops from three to two. The initial condition is set to be low flow. Later on, a surge occurs in the inflow demand, which forms the congestion when it reaches the lane-drop area. The congestion propagates backward until the demand surge ends and the inflow queue disappears, and then it discharges, leaving a triangular shape of low-speed area in the speed map.

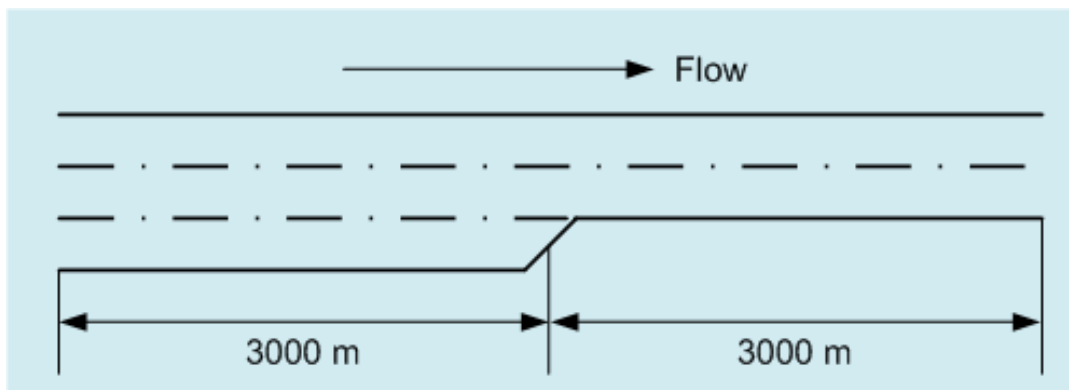


FIGURE 1 Experiment scenario.

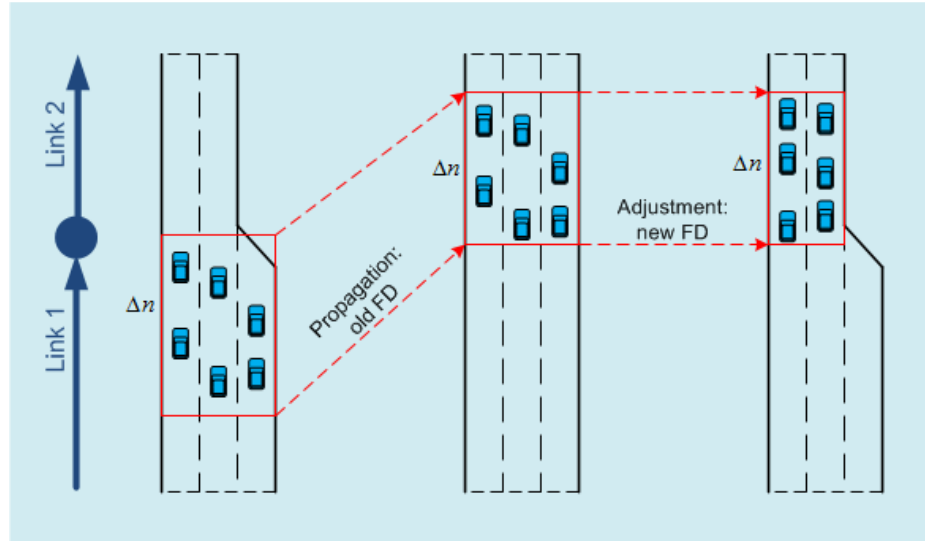


FIGURE 2 Lane-drop simulation model.

Lane-Drop Simulation Model

To realize the lane-drop scenario in discrete simulation, the simulation model explained in Figure 2 is adopted. When a vehicle group approaches the lane merging area, it is propagated ahead in the flow direction in an assumed three-lane freeway, and then the traffic characteristics (spacing, speed) are adjusted by assuming that the occupied area is the same while the number of lanes is reduced from three to two. The new spacing is calculated as $s_j^{k,new} = s_j^{k,old} \cdot (L_2 / L_1)$, where L_1 represents the number of lanes before the lane-drop, and L_2 represents the number of lanes after the lane-drop.

Experiment Settings

The Smulders' fundamental diagram is adopted, and the related parameters v_{cr} , v_f , s_{cr} , and s_{jam} are set to be 75 km/h, 120 km/h, 30 m/veh/lane, and 5 m/veh/lane. The simulation duration is 0.8 h to allow a full evolution of congestion, and the demand surge occurs between 0.1 and 0.3h. The low-demand input flow is 1 veh/s, while the high-demand input flow is 2veh/s. The time-step duration T is set to be 1 s, and the vehicle group size Δn is 4.6. For a first-order model, the CFL condition requires that $\Delta n > 2.5$ with $T = 1$. Both the first-order model and the second-order model are implemented in MATLAB. All these settings are the same for both models to make the outputs comparable.

The second-order model has three more parameters to calibrate, namely, τ , θ (or λ), and ε . Two nominal sets are tested, where Set 1 uses $\tau = 1.05$, $\theta = 0.55$, and $\varepsilon = 0.05$, and Set 2 uses $\tau = 1.14$, $\theta = 0.40$, and $\varepsilon = 0.05$. At this point, the parameter sets for the dynamic speed equation are chosen to ensure the simulation stability and yield reasonable (in the sense of traffic flow dynamics) outputs given other foregoing settings. They are not optimized towards a real data set. On the other hand, it is noticed that given other foregoing conditions, the parameters τ , θ , and ε can only vary within a small range while the simulation is still stable. Thus one may expect that

these parameters do not change greatly from the nominal data sets even when they are calibrated towards a real data set.

Outputs and Discussion

The simulation results are presented in Figure 3 to Figure 8. Figure 3 shows the speed map for the first-order model, with an amplified area shown in the Figure 4. Figure 5 and Figure 7 show the results for the second-order model with different parameter sets, with the same area amplified in Figure 6 and Figure 8, respectively. The lane-drop location and the amplified area are indicated in the figures. Initially, the demand volume is low such that the speed is close to the free flow speed, shown as the deep blue color. Between 360 s and 1,080 s, the demand volume is increased, resulting in a lower speed shown as light blue color before the lane-drop. As the plotting area is truncated at $x = 1,000$ m, this high-demand period is slightly shifted to the right in the figures. The increased inflow volume creates the congestion as it reaches the lane-drop location (the bottleneck), and the congestion propagates backward. When the inflow demand is decreased and the inflow queue is dissolved, the congestion begins to discharge.

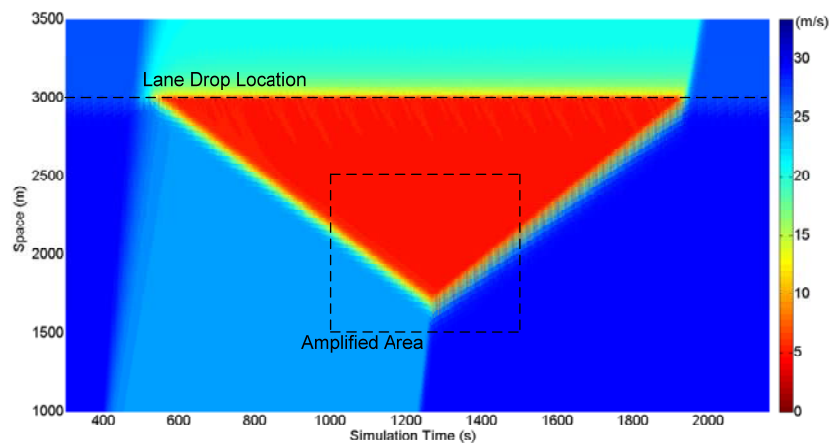


FIGURE 3 Speed map: first-order model.

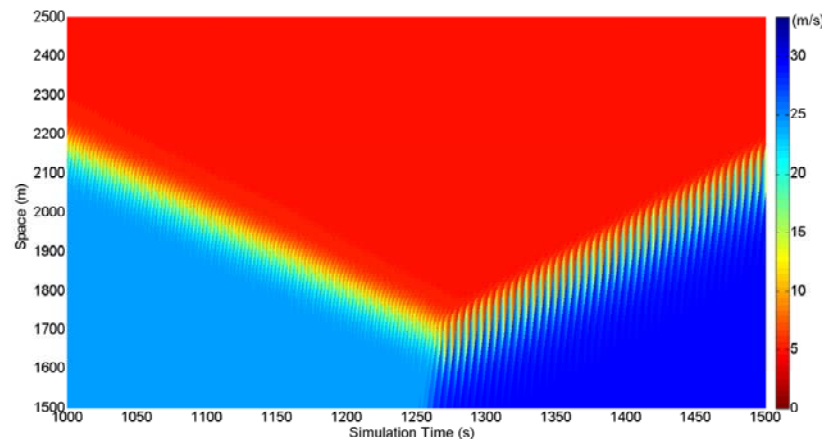


FIGURE 4 Speed map, amplified area: first-order model.

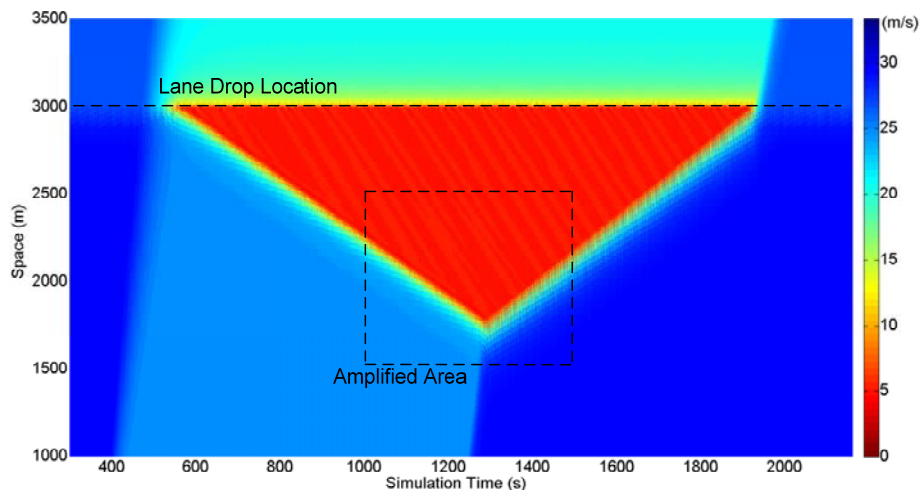


FIGURE 5 Speed map: second-order model with Parameter Set 1.

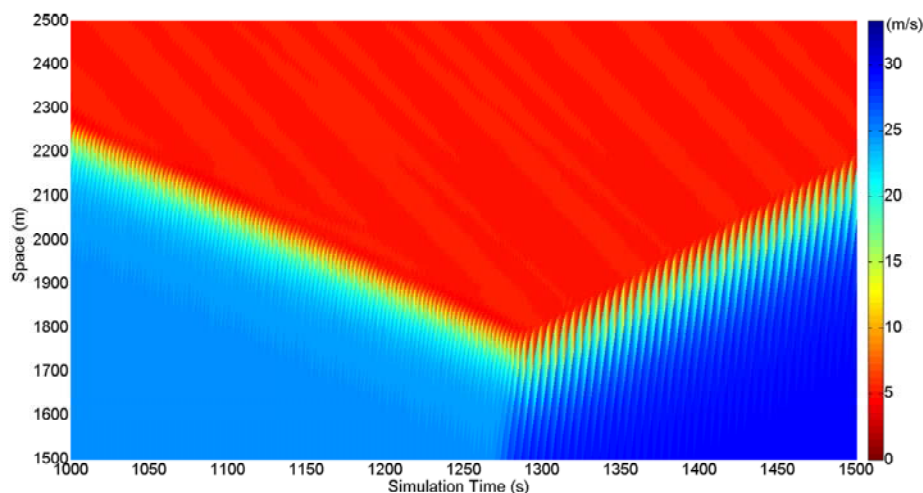


FIGURE 6 Speed map, amplified area: second-order model with Parameter Set 1.

We can see from the figures that all of them are able to reproduce the onset and discharge of the congestion caused by the lane-drop bottleneck. However, these figures have observable differences. Figure 5 exhibits similar pattern to that of the first-order model (Figure 3) but with some slight oscillations of speed, while Figure 7 shows more stop-and-go waves during the congestion. These stop-and-go waves are amplified towards the bottom of the congestion triangle. Researchers have found similar patterns in analyzing the real world data (17).

Van Wageningen-Kessels (2) has analyzed the stop-and-go waves that sometimes (e.g., low discretization resolution or CFL = 1) appear in the first-order Lagrangian model, and concluded that those waves are artificial as they are resulted from the diffusion error, and should not be confused with the real-world stop-and-go waves. However, the second-order model naturally creates the stop-and-go waves in congestion by adjusting how much the speed at current time-step relies on the speed during the last time-step, how fast it can reach the equilibrium speed, and how sensitive it is to the condition of the leading vehicles. When the

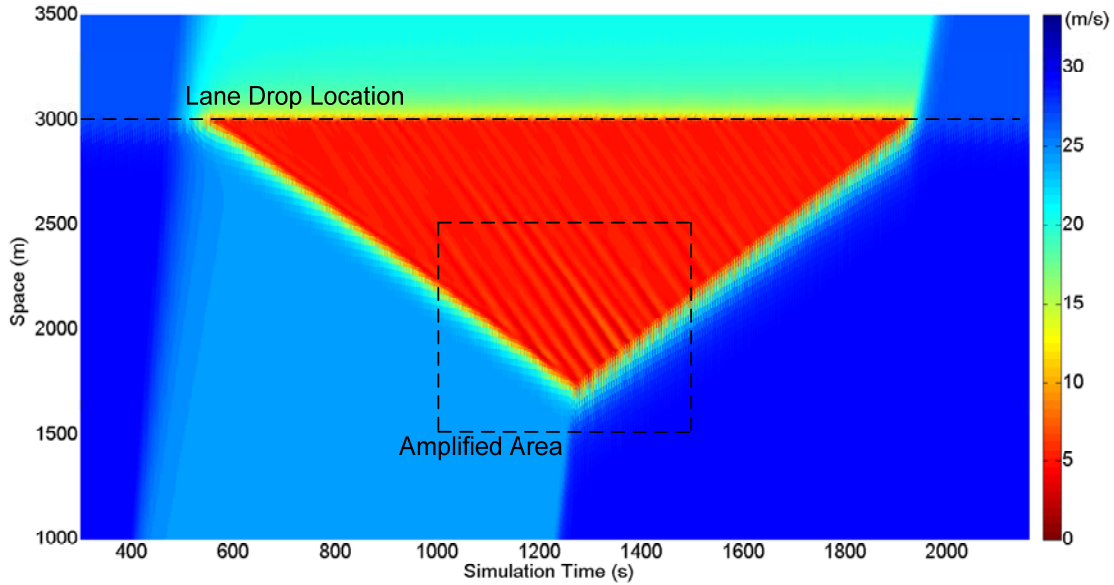


FIGURE 7 Speed map: second-order model with Parameter Set 2.

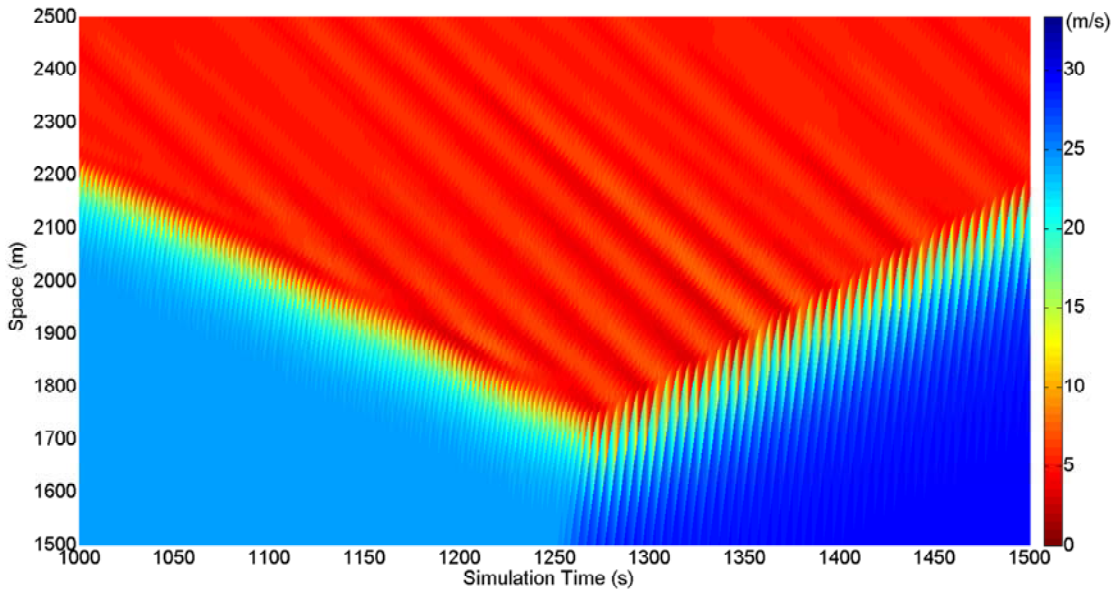


FIGURE 8 Speed map, amplified area: second-order model with Parameter Set 2.

reaction delay is relatively long, or that the vehicles are not sensitive to the front condition, the oscillation becomes more obvious (Parameter Set 2). In the contrary, when the drivers reach their desired speeds faster and are more sensitive to the front conditions (Parameter Set 1), the speed map becomes smoother. The resulted differences from the two parameter sets show the flexibility of the proposed model.

It is worth mentioning that the magnitude of the time delay parameter τ is one order higher than the drivers' reaction time in the microscopic level. In fact, this τ is two orders higher

than the drivers' reaction time in the Eulerian case. It should be treated as a macroscopic parameter, rather than being enforced to be the microscopic one.

CONCLUSIONS

In this paper, we propose a second-order Lagrangian macroscopic traffic flow model. A dynamic speed equation is derived from the realistic modeling of the reaction delay and driving anticipation. We show in a lane-drop simulation that the second-order model is not only able to reproduce the onset and discharge of congestion as the first-order model does, but also able to naturally create stop-and-go waves which are amplified as the waves propagate backward. These phenomena are coincident with the patterns identified in the real-world data. In addition, it can facilitate joint estimation of speed and incorporating speed measurements in traffic state estimation. All these observations show that the proposed model has advantages over the first-order model. However, there are still some fundamental topics that are not addressed. For example, no empirical validation has been conducted even for the first-order model. Calibration of model parameters would be another worthwhile future work. We are continuing our efforts in this direction.

REFERENCES

1. Leclercq, L., J. A. Laval, and E. Chevallier. The Lagrangian Coordinates and What It Means for First Order Traffic Flow Models. Presented at Transportation and Traffic Theory 2007, 2007.
2. Van Wageningen-Kessels, F., Y. Yuan, S. P. Hoogendoorn, H. Van Lint, and K. Vuik. Discontinuities in the Lagrangian Formulation of the Kinematic Wave Model. *Transportation Research Part C: Emerging Technologies*, Vol. 34, 2013, pp. 148–161.
3. Payne, H. J. Models of Freeway Traffic and Control. *Mathematical Models of Public Systems*, Simulation Councils, Inc., 1971.
4. Lighthill, M. J., and G. B. Whitham. On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads. *Proc., the Royal Society of London. Series A. Mathematical and Physical Sciences*, Vol. 229, No. 1178, 1955, pp. 317–345.
5. Richards, P. I. Shock Waves on the Highway. *Operations Research*, Vol. 4, No. 1, 1956, pp. 42–51.
6. Daganzo, C. F. The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research Part B: Methodological*, Vol. 28, No. 4, 1994, pp. 269–287.
7. Daganzo, C. F. The Cell Transmission Model, Part II: Network Traffic. *Transportation Research Part B: Methodological*, Vol. 29, No. 2, 1995, pp. 79–93.
8. Newell, G. F. A Simplified Theory of Kinematic Waves in Highway Traffic, Part I: General Theory. *Transportation Research Part B: Methodological*, Vol. 27, No. 4, 1993, pp. 281–287.
9. Papageorgiou, M., J. M. Blosseville, and H. Hadj-Salem. Macroscopic Modelling of Traffic Flow on the Boulevard Périphérique in Paris. *Transportation Research Part B: Methodological*, Vol. 23, No. 1, 1989, pp. 29–47.
10. Wang, Y., and M. Papageorgiou. Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter: A General Approach. *Transportation Research Part B: Methodological*, Vol. 39, No. 2, 2005, pp. 141–167.
11. Daganzo, C. F. Requiem for Second-Order Fluid Approximations of Traffic Flow. *Transportation Research Part B: Methodological*, Vol. 29, No. 4, 1995, pp. 277–286.

12. Pavari-Fontana, S. L. On Boltzmann-Like Treatments for Traffic Flow: A Critical Review of the Basic Model and an Alternative Proposal for Dilute Traffic Analysis. *Transportation Research*, Vol. 9, No. 4, 1975, pp. 225–235.
13. Papageorgiou, M. Some Remarks on Macroscopic Traffic Flow Modelling. *Transportation Research Part A: Policy and Practice*, Vol. 32, No. 5, 1998, pp. 323–329.
14. Smulders, S. Control of Freeway Traffic Flow. Ph.D. dissertation. University of Twente, Enschede, Netherlands, 1989.
15. Yuan, Y., J. W. C. Van Lint, R. E. Wilson, F. van Wageningen-Kessels, and S. P. Hoogendoorn. Real-Time Lagrangian Traffic State Estimator for Freeways. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 13, No. 1, 2012, pp. 59–70.
16. Lebacque, J. P. The Godunov Scheme and What It Means for First-Order Traffic Flow Models. *International Symposium on Transportation and Traffic Theory*, 1996, pp. 647–677.
17. Zielke, B. A., R. L. Bertini, and M. Treiber. Empirical Measurement of Freeway Oscillation Characteristics: An International Comparison. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2088, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 57–67.

APPENDIX A
Keynote Speakers'
Powerpoint Presentations

KEYNOTE ADDRESS

Reflections on 50 Years of the TRB Committee on Traffic Flow Theory and Characteristics

NATHAN H. GARTNER
University of Massachusetts

KEYNOTE ADDRESS

Reflections on 50 Years of the TRB Committee on Traffic Flow Theory and Characteristics


Prof. Nathan H. Gartner
University of Massachusetts
Lowell, MA, USA
© NHGartner, UML, August 2014



Learning with Purpose

AHB45
Committee on Traffic Flow Theory & Characteristics


Transportation Research Board
National Academy of Sciences



Learning with Purpose

History

Highway Research Board
Department of Traffic and Operations
1956
Wilbur S. Smith
Chairman



Learning with Purpose

History – Predecessor Committees 50 years ?

COMMITTEE CHAIRS	AFFILIATION	YEARS SERVED
COMMITTEE ON TRAFFIC FLOW THEORY (1963-1970) TO-9		
D.L. Gerlough	Planning Research Corp.	1963
Donald E. Cleveland	University of Virginia	1964-1967
COMMITTEE ON CHARACTERISTICS OF TRAFFIC FLOW (1963-1970) TO-12		
Harold L. Michael	Purdue University	1963-1964
Joseph C. Oppenlander	University of Vermont	1964-1968
Donald C. Capelle	Alan M. Voorhees & Assoc. Inc.	1968-1970



Learning with Purpose

Predecessor Committees

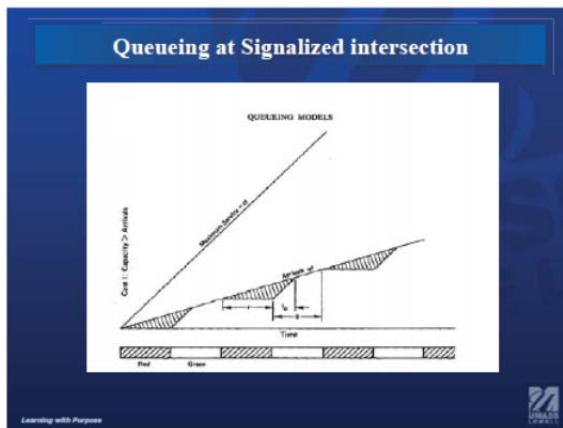
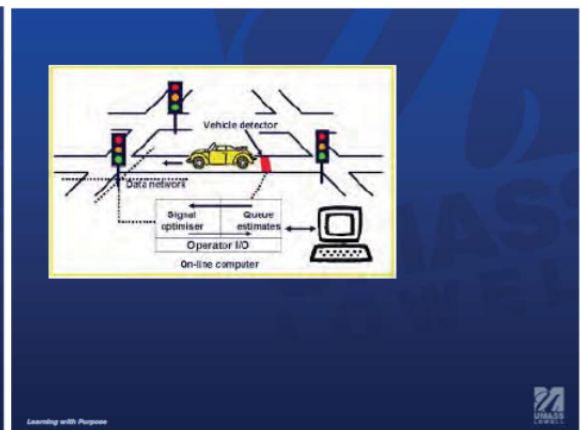
COMMITTEE CHAIRS	AFFILIATION	YEARS SERVED
COMMITTEE ON SPEED CHARACTERISTICS (1939-1963) TO-13A		
Burton W. Marsh	AAA	1939-1954
Harold L. Michael	Purdue University	1955-1963

Project Committees (1956)

COMMITTEE NAMES
Parking
Origin and Destination Surveys
Influence of Shoulders on Traffic Operations
Highway Capacity
Traffic Control Devices
Vehicle Characteristics
Operating Speeds in Urban Areas
Roadway Pavement Markings
Channelization
Speed Characteristics
Urban Volume Characteristics
Road User Characteristics
Ad Hoc Committee on Traffic Flow Theory
Ad Hoc Committee on Urban Congestion Index
Ad Hoc Committee on Highway Safety

Ad Hoc Committee on Traffic Flow Theory (1956)

OFFICERS		
J. H. Mathewson		Chairman
Willard Bascom		
H. H. Goode		
Bruce D. Greenshields		
Eugene Maier		
William Prager		
C. W. Prisk		
Deane S. Terry, Jr.		



Committee on Traffic Flow Theory and Characteristics (A3A18 & AHB45)

COMMITTEE CHAIRS	AFFILIATION	YEARS SERVED
Donald C. Capelle	Alan M. Voorhees & Assoc. Inc.	1971-1974
Kenneth W. Crowley	Pennsylvania State University	1974-1980
John J. Haynes	University of Texas at Arlington	1980-1986
Carrol J. Messer	Texas A&M University	1986-1992
Hani S. Mahmassani	University of Texas at Austin	1992-2000
Nathan H. Gartner	University of Massachusetts Lowell	2000-2009
Robert L. Bertini	Portland State University	2009-2015

Source: <http://www.ttpu.edu/members.htm>

Committee on Traffic Flow Theory & Characteristics (AHB45)

COMMITTEE CHAIRS	AFFILIATION	YEARS SERVED	TRB STAFF
Donald C. Capelle	Alan M. Voorhees & Assoc. Inc.	1971-1974	K.B. Johns
Kenneth W. Crowley	Pennsylvania State University	1974-1980	K.B. Johns David K. Witthoford
John J. Haynes	University of Texas at Arlington	1980-1986	David K. Witthoford
Carrol J. Messer	Texas A&M University	1986-1992	Richard A. Cunard
Hani S. Mahmassani	University of Texas at Austin	1992-2000	Richard A. Cunard
Nathan H. Gartner	University of Massachusetts Lowell	2000-2009	Richard A. Cunard
Robert L. Bertini	Portland State University	2009-2015	Richard A. Cunard

source: <http://www.tl.fh.edu/members.htm>

Monographs on Traffic Flow Theory

I – 1964
II – 1975
III – 2001

MONOGRAPH I

An Introduction to Traffic Flow Theory

HRB, 1964

HIGHWAY RESEARCH BOARD
Special Report 79

AN INTRODUCTION TO Traffic Flow Theory

Edited by
 D.L. Gerlough
 and
 D.G. Capelle

HIGHWAY RESEARCH BOARD
 of the
 Division of Engineering and Industrial Research
 National Academy of Sciences
 National Research Council
 Washington, D.C.
 1964

Highway Research Board

Officers and Members of the Executive Committee
1964

Highway Research Board (1964)

OFFICERS	
NAME	DESIGNATION
Wilbur S. Smith	Chairman
Donald S. Berry	First Vice Chairman
J. B. McMorran	Second Vice Chairman
D. Grant Mickle	Executive Director
W.N. Carey, Jr.	Deputy Executive Director
Fred Burggraf	Consultant

EDITORIAL STAFF	
D. Grant Mickle	
Herbert P. Orland	
Earle W. Jackson	

Highway Research Board (1964)

EXECUTIVE COMMITTEE	
NAME	AFFILIATION
Rex M. Whitton	Bureau of Public Roads
A.E. Johnson	American Association
Louis Jordan	National Research Council
R.R. Bartelsmeyer	H.W. Lochner & Co.
C.D. Curtiss	American Road Builders' Association
E.W. Bauman	National Slag Association
Donald S. Berry	Northwestern University
W.A. Bugge	Parsons Brinckerhoff-Tudor-Bechtel
Mason A. Butcher	Montgomery County, Md.
J. Douglas Carroll, Jr.	Tri-State Transportation Committee, NYC
Harmer E. Davis	University of California
Duke W. Dunbar	Attorney General of Colorado
John T. Howard	Massachusetts Institute of Technology

Highway Research Board (1964)

EXECUTIVE COMMITTEE	
NAME	AFFILIATION
Pyke Johnson	Retired
Louis C. Lundstorm	General Motors Proving Grounds
Burton W. Marsh	American Automobile Association
Oscar T. Marzke	U.S. Steel Corporation
J.B. McMorran	New York State Dept. of Public Works
Clifford S. Rassweiler	Johns-Manville Corporation
M.L. Shadburn	Georgia State Highway Dept.
T.E. Shelburne	Virginia Dept. of Highways
Wilbur S. Smith	Wilbur Smith Associates
John H. Swanberg	Minnesota Dept. of Highways
Edward G. Wetzel	The Port of New York Authority
K.B. Woods	Purdue University

Highway Research Board (1964)

SPECIAL COMMITTEE ON PUBLICATION OF SELECTED INFORMATION ON THEORY OF TRAFFIC FLOW	
Representatives of the Highway Research Board	
	D. E. Cleveland
	R. S. Foote
	D. L. Gerlough
	C. C. Robinson
Representatives of the Division of Mathematics	
	H. P. Galligher
	E. W. Montroll
	G. F. Newell
	R. W. Rothery

Highway Research Board (1964)

CONTENTS		
CHAPTER TITLE	AUTHORS	AUTHOR AFFILIATION
1. Hydrodynamic Approaches Part I Part II	L. A. Pipes M. J. Lighthill G. B. Whitman	University of California Royal Aircraft Establishment California Institute of Technology
2. Car Following and Acceleration Noise	E. W. Montroll R. B. Potts	International Business Machines Corporation University of Adelaide, Australia
3. Queueing Theory Approaches	D. E. Cleveland D. G. Capelle	Texas A&M University Automotive Safety Foundation
4. Simulation of Traffic Flow	D. L. Gerlough	Planning Research Corporation
5. Some Experiments and Applications	R. S. Foote	The Port of New York Authority

MONOGRAPH II

Traffic Flow Theory A Monograph

TRB, 1975

Traffic Flow Theory

A Monograph

PERC
SPECIAL REPORT 165
Transportation Research Board
National Research Council

TRANSPORTATION RESEARCH BOARD

Officers and Members of the Executive Committee 1975




Transportation Research Board (1975)

AUTHORS	
Daniel L. Gerlough	
Matthew J. Huber	


EDITOR	
H.P. Orland	

OFFICERS	
NAME	DESIGNATION
Milton Pikarsky	Chairman
Harold L. Michael	Vice Chairman
W. N. Carey, Jr.	Executive Director



Transportation Research Board (1975)

EXECUTIVE COMMITTEE	
NAME	AFFILIATION
Harvey Brooks	National Research Council
Asaph H. Hall	Federal Railroad Administrator
Robert E. Patricelli	Urban Mass Transportation Administrator
Henrik E. Stafseth	American Association of State Highway and Transportation Officials
Norbert T. Tiemann	Federal Highway Administrator
William L. Garrison	University of California Berkeley
Jay W. Brown	Florida Department of Transportation
George H. Andrews	Sverdrup and Parcel
Kurt W. Bauer	Southeastern Wisconsin Regional Planning Commission
Manuel Carballo	Wisconsin Department of Health and Social Sciences
L. S. Crane	Southern Railway Systems
James M. Davey	Consultant, MI




Transportation Research Board (1975)

EXECUTIVE COMMITTEE	
NAME	AFFILIATION
Howard L. Gauthier	Ohio State University
Alfred Hedefine	Parsons, Brinckerhoff, Quade and Douglas Inc.
Robert N. Hunter	Missouri State Highway Commission
A. Scheffer Lang	Association of American Railroads
Benjamin Lax	Massachusetts Institute of Technology
Daniel McFadden	University of California Berkeley
Harold L. Michael	Purdue University
D. Grant Mickle	Highway Users Federation for Safety and Mobility
James A. Moe	Bechtel Inc.
Milton Pikarsky	Regional Transportation Authority, Chicago




Transportation Research Board (1975)

EXECUTIVE COMMITTEE	
NAME	AFFILIATION
J. Philip Richley	Transportation, Dalton, Dalton, Little and Newport
Raymond T. Schuler	New York State Department of Transportation
William K. Smith	General Mills Inc.
B. R. Stokes	American public transit Association
Percy A. Wood	United Air Lines
Louis J. Gambaccini	Port Authority Trans-Hudson Corporation



Transportation Research Board (1975)

ADVISORY COMMITTEE	
NAME	AFFILIATION
Robert S. Foote	Port Authority of New York and New Jersey
Dr. Adolf D. May, Jr.	University of California
Richard W. Rothery	General Motors Research Laboratories
K. B. Johns	TRB Liaison Representative
Sidney Weiner	Federal Highway Administration
Barry Benioff	Federal Highway Administration
William W. Wolman	Federal Highway Administration
Donald G. Capelle	Alan M. Vorhees & Associates



Transportation Research Board (1975)

CONTENTS	
CHAPTER NUMBER	CHAPTER TITLE
1.	Introduction
2.	Measurement of Flow, Speed and Concentration
3.	Statistical Distributions of Traffic Characteristics
4.	Traffic Stream Models
5.	Driver Information Processing Characteristics
6.	Car Following and Acceleration Noise
7.	Hydrodynamic and Kinematic Models of Traffic
8.	Queueing Models
9.	Simulation of Traffic Flow
10.	Epilogue

MONOGRAPH III

Traffic Flow Theory A State-of-the-Art Report

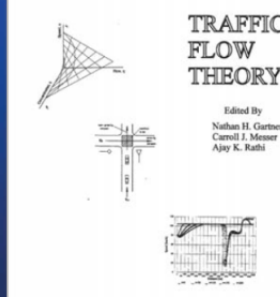
TRB-FHWA-ORNL
2001

Traffic Flow Theory A State-of-the-Art Report

Revised
2001

TRAFFIC FLOW THEORY

Edited By
Nathan H. Gartner
Carroll J. Messer
Ajay K. Rathi



U.S. Department of Transportation
Federal Highway Administration

Oak Ridge National Laboratory
omi

Transportation Research Board

Officers and Members of the Executive Committee

2001

Transportation Research Board (2001)

EDITORS	
NAME	AFFILIATION
Dr. Nathan Gartner	University of Massachusetts - Lowell
Dr. Carrol J. Messer	Texas A & M University
Dr. Ajay K. Rathi	Oak Ridge National Laboratory

ADVISORY COMMITTEE	
NAME	DESIGNATION
Richard Cunard	TRB Liaison Representative
Dr. Henry Lieu	Federal Highway Administration
Dr. Hani Mahmassani	University of Texas at Austin

Transportation Research Board (2001)

REVIEW COMMITTEE		
NAME	NAME	NAME
Rahmi Akcelik	Mike Florian	John Leonard II
Rahim Benekohal	Fred Hall	Fred Mannering
David Boyce	Benjamin Heydecker	William McShane
Michael Brackstone	Ben Hurdle	Kyriacos Mouskos
Werner Brilon	Shinya Kikuchi	Vladimir Protopopescu
Christine Buisson	Helmut "Bill" Knee	Tom Rockwell
Ennio Cascetta	Haris Koutsopolous	Natacha Thomas
Michael Cassidy	Jean-Baptiste Lesort	Ken Courage
Avishai Cedar	Arun Chatterjee	Rey Derr
Panos Prevedourous	Bin Ran	Mitsuru Saito

Transportation Research Board (2001)

CONTENTS		
CHAPTER TITLE	AUTHOR	AUTHOR AFFILIATION
1. Introduction	Nathan H. Gartner Carrol Messer Ajay K. Rathi	University of Massachusetts Lowell Texas A&M University Oak Ridge National Laboratory
2. Traffic Stream Characteristics	Fred L. Hall	McMaster University
3. Human Factors	Rodger J. Koppa	Texas A&M University
4. Car Following Models	Richard W. Rothery	The University of Texas at Austin
5. Continuum Flow Models	H. Michael Zhang Reinhart Kuhne Panos Michalopolous	University of California Davis Steierwald Schonharting und Partner University of Minnesota Institute of Technology
6. Macroscopic Flow Models	James C. Williams	University of Texas at Arlington

Transportation Research Board (2001)

CONTENTS		
CHAPTER TITLE	AUTHOR	AUTHOR AFFILIATION
7. Traffic Impact Models	Sia Ardekani Ezra Hauer Bahram Jamei	University of Texas at Arlington University of Toronto Virginia Department of Transportation
8. Unsignalized Intersection Theory	Rod J. Troutbeck Werner Brilon	Queensland University of Technology Ruhr University
9. Traffic Flow at Signalized Intersections	Nagui Rouphail Andrzej tarko Jing Li	North Carolina State University Purdue University TransSmart Technologies
10. Traffic Simulation	Edward Lieberman Ajay K. Rathi	KLD Associates Oak Ridge National Laboratory
11. Kinetic Theories	Paul Nelson	Texas A&M University

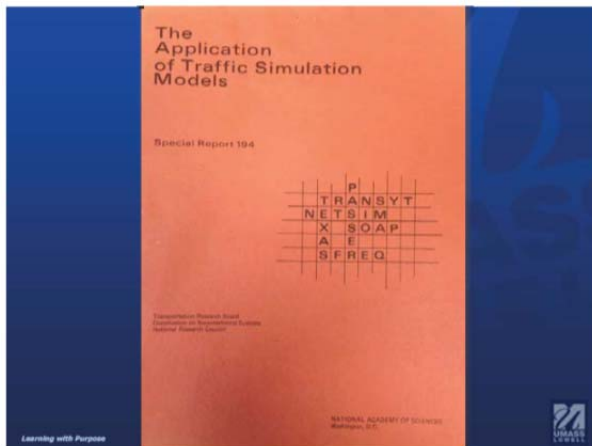
Summer Meetings TFTC Committee

2008 – Woods Hole, MA
2010 – Annecy, France
2012 – Ft. Lauderdale, FL
2014 – Portland, ORL

1981 – Williamsburg, VA
1996 – ISTTT13, Lyon, France

Summer Meetings TFTC Committee

2008 – Woods Hole, MA
2010 – Annecy, France
2012 – Ft. Lauderdale, FL
2014 – Portland, OR



The Application of Traffic Simulation Models
Special Report 194
Transportation Research Board
Commission on Sociotechnical Systems
National Research Council.

Steering Committee for the Conference on the Application of Traffic Simulation Models

Chairman

Donald E. Orne, Michigan Department of Transportation

Members

Kenneth G. Courage, University of Florida
Kenneth W. Crowley, Institute for Research
Nathan H. Gartner, University of Lowell
Herman E. Haenel, Texas State Department of Highways and Public Transportation
Joseph K. Lam, DeLeuw, Cather, Canada
Edward B. Lieberman, KLD Associates, Inc.
Paul Ross, Federal Highway Administration

Liaison Representatives

David R.P. Gibson and H. Milton Heywood, Federal Highway Administration

Transportation Research Board Staff

David K. Witheford

Nathan H. Gartner, Conference Proceedings Editorial Consultant

SIMULATION SOFTWARE

NAME	DATE	APPLICATION	MODELLING APPROACH	PROGRAM LANGUAGE	COMPUTER
TEXAS	1978	Traffic Performance	Mic., Stoc., TS, Sim.	Fortran IV	CDC 6600 IBM 370
SOAP	1977	Signal Timing (Cycle splits & phasing)	Mac., Det., TS, Opt.	Fortran IV	IBM 360/370
UTCS-1	1973	Traffic Performance	Mic., Stoc., TS, Sim.	Fortran IV	IBM 360
TRANSYT7	1978	Opt. Signal Timing	Mac., Det., TS, Opt.	Fortran IV	ICL 4-70 IBM 360/370
NETSIM	1977	Evaluate Signal Control Systems	Mic., Stoc., TS, Sim.	Fortran IV	IBM 360/370 CDC 6600
FREQ(N) N = 3,4,...,12	1975	Develop Optimal Ramp Metering, Evaluate HOV Lanes	Mac., Det., TS, Opt.	ANSI Fortran, Fortran IV	CDC/IBM

Abbreviations:
 Mic. - Microscopic
 Deterministic
 TS - Time Scan
 Sim. - Simulation
 Mac. - Macroscopic
 Stoc. - Stochastic
 ES - Event Scan
 Opt. - Optimization
 Det. -

SIMULATION SOFTWARE

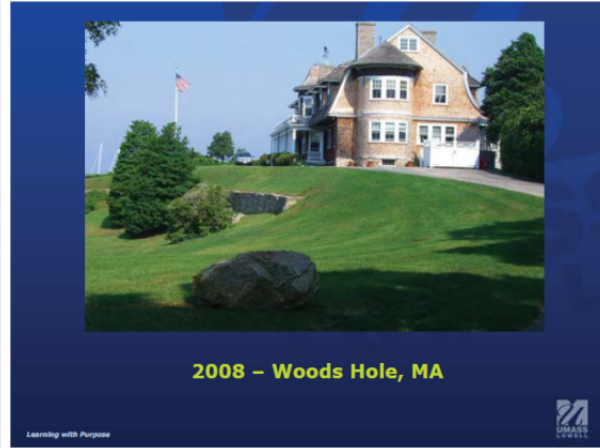
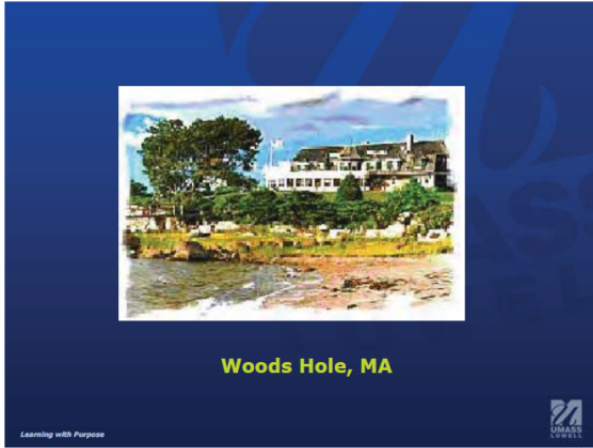
NAME	DATE	APPLICATION	MODELLING APPROACH	PROGRAM LANGUAGE	COMPUTER
INTRAS	1977	Eva. Freeway incidents on Corridor Operations	Mic., Stoc., TS, Sim.	Fortran IV	IBM 370 CDC 7600
CORQ	1974	Eva. Traffic Control Strategies within Corridor	Mic., Det., TS, Sim.	Fortran IV	Unknown
STAR	1974	Eva. Surveillance and Control Strategies for Route Diversion	Mac., Det., TS, Sim.	Unknown	Unknown
SCOT	1975	Eva. Traffic Control Strategies within Corridor	Mac., Det., TS, Sim.	Fortran IV	CDC 660 IBM 370 UNIVAC
TRANSIM	1966	Eva. of Traffic Performance in System	Mic./Mac., Stoc./Det., TS, Sim.	Fortran IV	IBM 7090 7094, 1401

Abbreviations:
 Mic. - Microscopic
 Deterministic
 TS - Time Scan
 Sim. - Simulation
 Mac. - Macroscopic
 Stoc. - Stochastic
 ES - Event Scan
 Opt. - Optimization
 Det. -

TRANSPORTATION RESEARCH
CIRCULAR
Number E-C149 June 2011


75 Years of the Fundamental Diagram for Traffic Flow Theory

Greenshields Symposium





Foundations of Traffic Flow Theory The Fundamental Diagram

Greenshields, B. D.
A Study of Traffic Capacity.
Highway Research Board Proceedings.
Vol. 14, 1935, pp. 448-477.



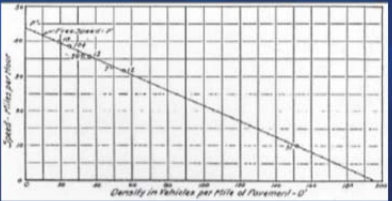
Learning with Purpose

Greenshields Measuring Traffic in Ohio, 1930s





Learning with Purpose

The Linear Speed-Density Relation

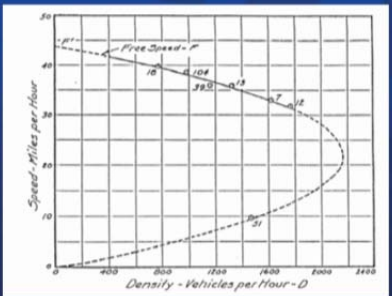


Source: [Greenshields, 1934].




Learning with Purpose

The First Fundamental Diagram as a q-v Relationship

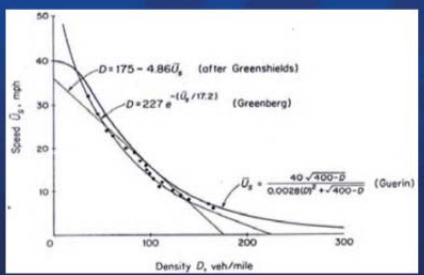



Source: [Greenshields, 1934].



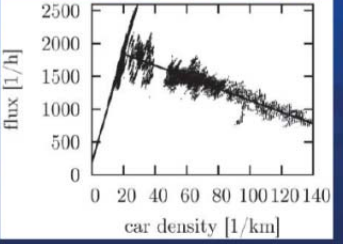
Learning with Purpose

Various Density-Speed Relationships





Learning with Purpose

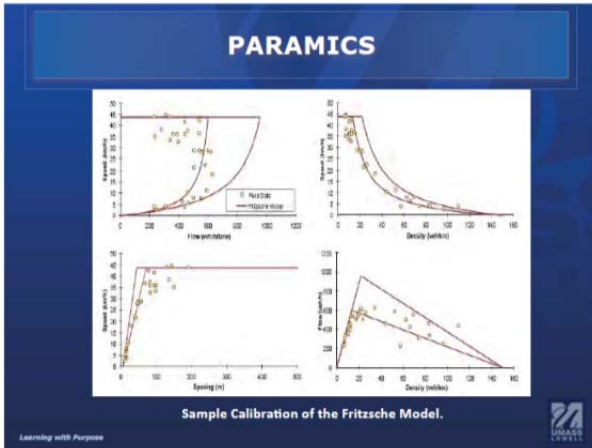
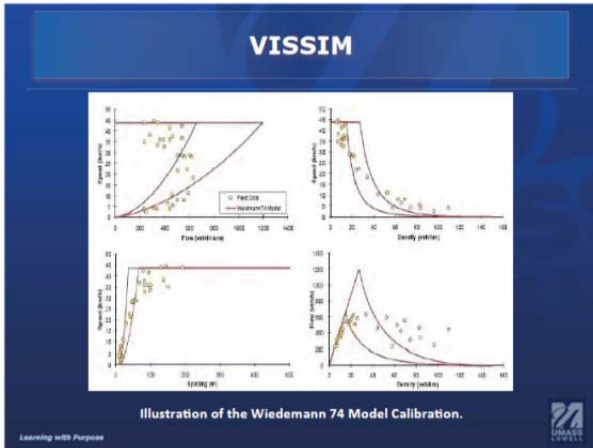
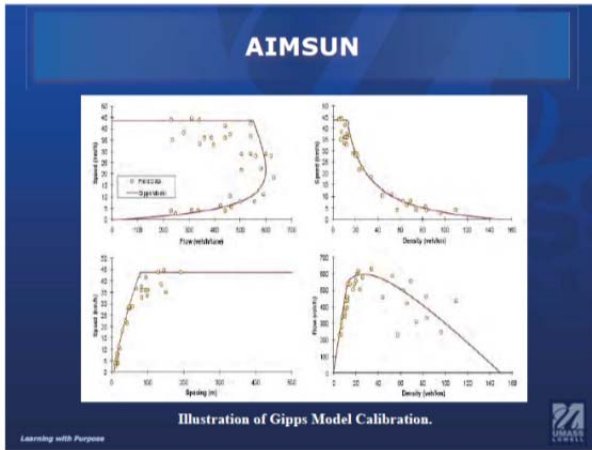
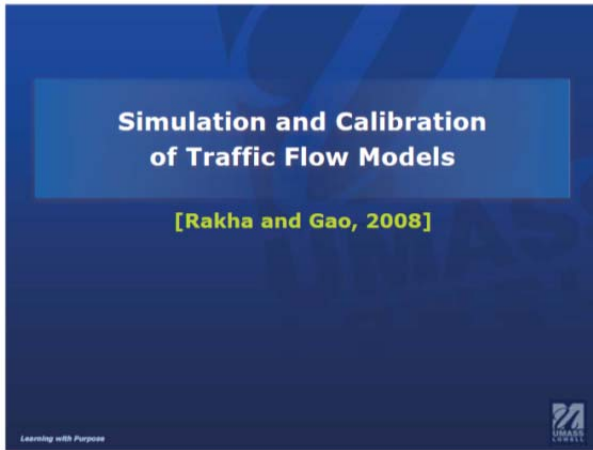
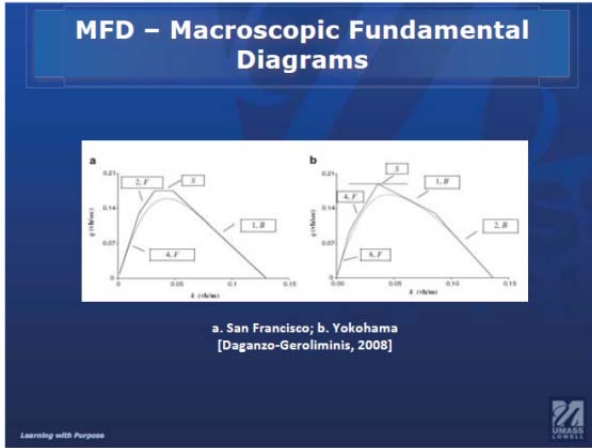
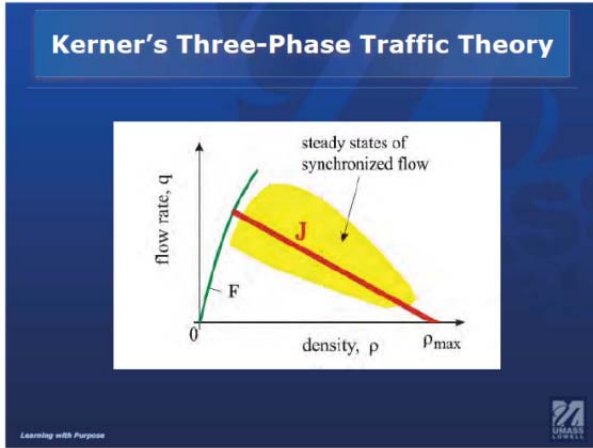
Fundamental Diagram of Real Traffic Data

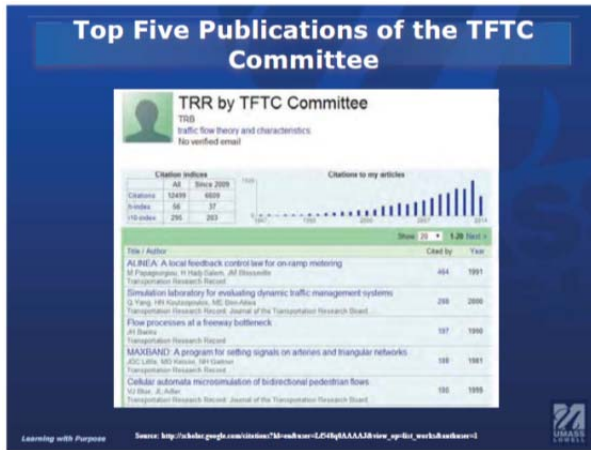
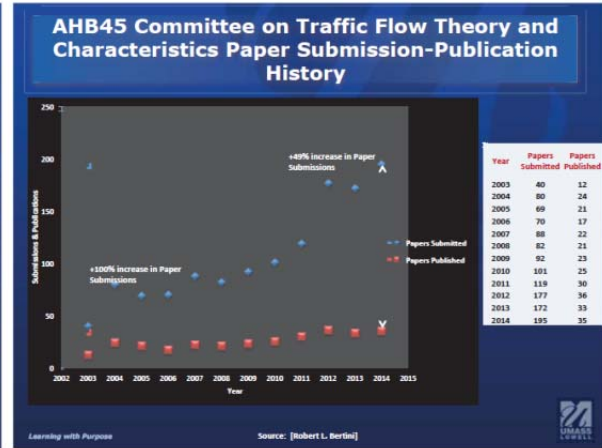
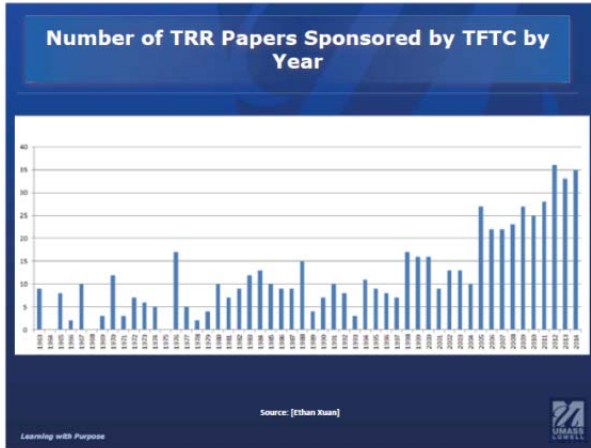


Linear Fitting of the Two Traffic Regions.



Learning with Purpose





AHB45 – Committee on Traffic Flow Theory and Characteristics

"Quo vadis?"
"Romam vado iterum crucifigi"

75 YEARS – 2039

100 YEARS – 2064

THE END

QUESTIONS?

KEYNOTE ADDRESS

Exploring the Impact of Microscopic Features of Traffic on Macroscopic Patterns

VINCENT PUNZO

University of Naples Federico II, Italy



Contributors



Exploring the impact of microscopic features of traffic on macroscopic patterns

Vincenzo Punzo

University of Naples Federico II, Italy

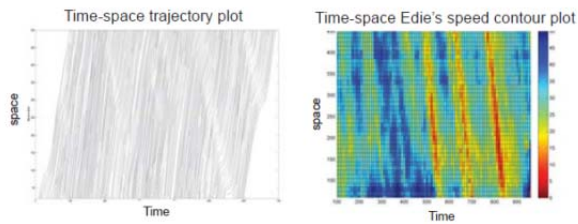
- **Marcello Montanino**
- **Biagio Ciuffo** (now at the European Commission Joint Research Centre)

Traffic Flow Theory and Characteristics Committee (AHB45)
 2014 Summer Meeting - August 11-13, 2014 - Portland, Oregon USA
 Symposium Celebrating 50 Years of Traffic Flow Theory

The twofold representation of traffic provided by NGSIM(-like) data



- NGSIM: trajectory of each single vehicle crossing the observed time-space domain → *spatiotemporal traffic patterns from microscopic data*
- Other traffic data: very partial view of the traffic phenomenon



- The traffic scientific community has not yet thoroughly exploited this twofold 'value' of NGSIM data

Twofold representation and TFT studies (1)



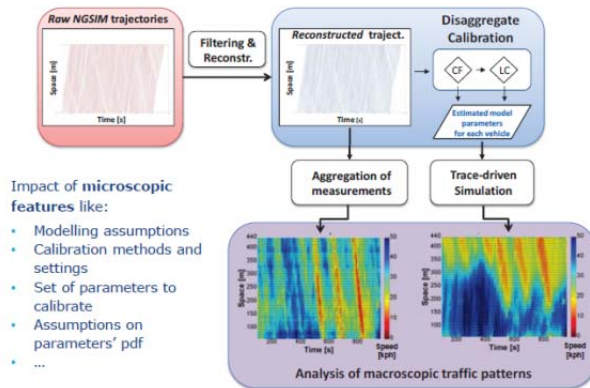
In microscopic modelling of traffic:

- We rather know how each single component works alone:
 - ✓ Theoretical investigations and empirical analyses of single 'driver' models (CF or LC models)
 - ✓ In car-following, most often single-lane, homogeneous-flow assumptions
 - ✓ Calibration and validation on disaggregate data focused on reproducing the single behaviour
- We do not really know what happens when single components interact in a stochastic traffic simulation environment:
 - ✓ Interaction effect of LC and CF decisions (i.e. of models output)?
 - ✓ Impact of parameter heterogeneity?
 - ✓ Impact of assumptions on the probabilistic model of parametric inputs?

Twofold representation and TFT studies (2)

- What is the impact of microscopic features and assumptions on the collective behaviour of traffic? I.e. as captured by spatiotemporal congested patterns?
- ✓ Spatiotemporal patterns are not the most often used measure to investigate microscopic models (often in simple settings like e.g. car-following only; Treiber & Kesting, 2012, *TR-C*)
- **Dichotomy** between microscopic TFT and traffic micro-simulation
- ✓ Traffic simulation outcomes mostly investigated by means of time series or frequency plots

Analysis framework



Methodology in extreme synthesis

- Using **measured and simulated macroscopic spatiotemporal traffic patterns** to investigate impact of **microscopic features**
- Calibrating on disaggregate data and validating on collective data

Calibration of IDM and MOBIL

(Treiber et al., 2000, *Ph. Rev. E*; Kesting et al., 2007, *TRR*)

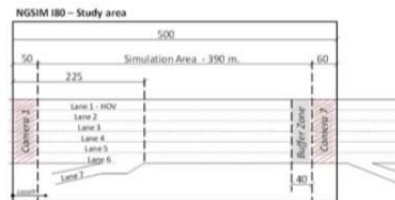
- **IDM**:
 - ✓ $m \ln(\text{RMSE}(\text{spacing}))$
 - ✓ OptQuest Multistart (Punzo et al. 2012, *TRR*)
 - ✓ Uncertainty analysis and physical informed criteria to set parameters bounds (Punzo et al. 2014, *IEEE T-ITS*)
- **MOBIL** (no calibration attempts in literature; (*) Zheng, 2014, *TR-B*):
 - ✓ The lane-changing event is rare → not interested in the instant and its traffic conditions but in the prevailing traffic conditions that generate it
 - ✓ scenario → leaders & followers in current and target lanes do not change
 - ✓ $m \ln \left\{ \frac{\#LC}{\sigma_{LC}} + \frac{\#noLC}{\sigma_{noLC}} \right\}$; where:
 - #LC = number of unsuccessful LC scenarios (*1-detection rate*^(*))
 - #noLC = number of unsuccessful noLC scenarios (*false alarm rate*^(*))
 - $\sigma_{LC} \in \sigma_{noLC}$ computed through Monte Carlo uncertainty analysis

Calibration of MOBIL (2)

- ✓ MOBIL LC model calculates the potential advantage of all the vehicles involved in the LC maneuver in terms of acceleration, as given by the IDM model;
- ✓ MOBIL calibration is therefore conditional on the IDM calibration;
- ✓ In calibration:
 - ✓ vehicles are moved according to the NGSIM measurements (and not using the IDM).
 - ✓ The IDM is used to calculate accelerations yielding the potential advantage, by using for each vehicle its own IDM calibrated parameters;

Trace driven micro-simulation (NGSIM I80)

- **Trace-driven** traffic micro-simulation → *fair comparison*
- Vehicle insertion as in measured data
- Downstream conditions superimposed to exiting vehicles
- ...
- **Car-following: IDM**
- **Lane Changing: MOBIL**
- **Merging: mandatory lane-changing with MOBIL**



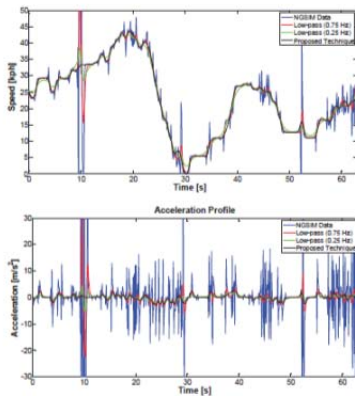
NGSIM DATA ERROR ANALYSIS AND RECONSTRUCTION

Opened issues

- To which analyses can NGSIM data be reliably applied to without any processing?
- Which is the impact of such errors on analyses made on NGSIM data?
- Which the accuracy requirements for future data gathering?
- ...
- Since 2012, 5 out of 19 studies published on journals^(*) using NGSIM trajectories applied some kind of filter to data:
 - ✓ 2 out of 11 on car-following
 - ✓ 3 out of 8 on lane-changing

(*) source Scopus, accessed 28/07/2014

Inadequateness of low-pass filters



Background on analysis of NGSIM data

Thieman, Treiber & Kesting, 2008, *TRR*

- Because of the noise in the positional data, velocity and acceleration information cannot be extracted directly
- Symmetric exponential moving average filter to be applied "to all trajectories before any further data analysis"

Punzo, Borzacchiello & Ciuffo, 2009, *TRB*; 2011, *TR-C*

- General methodology to quantify the degree of accuracy/bias in vehicle trajectory data
- Different criteria:
 - ✓ Analysis of accelerations, jerks, amplitude frequency spectrum
 - ✓ Internal consistency, platoon consistency
 - ✓ Application to all the NGSIM datasets
 - High level of measurement errors
 - 4.0-12.4% of leader-follower couples show unphysical inter-vehicle spacing

Filtering NGSIM data

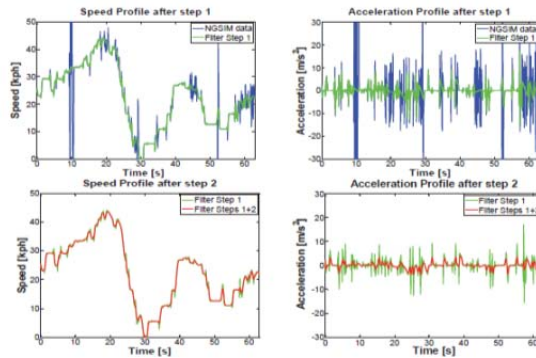
- Usual filtering techniques are inadequate (e.g. kernel smoothing)
- If you cannot filter the data, just throw them away...
- If the time window of eliminated points is short, lots of available information/constraints to reconstruct the missing trajectory:
 - ✓ space travelled within the window;
 - ✓ physical capabilities of cars;
 - ✓ car-following dynamics (inter-vehicle spacing integrity).

➔ Multi-step reconstruction procedure

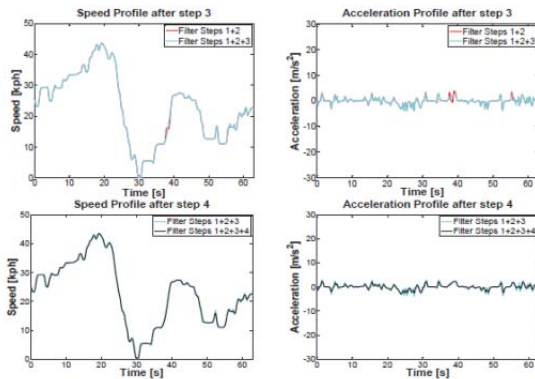
Montanino and Punzo 2011, *TRR*
 Montanino and Punzo 2015, *TRB*

Application to I80-1 dataset (available at www.multitude-project.eu)

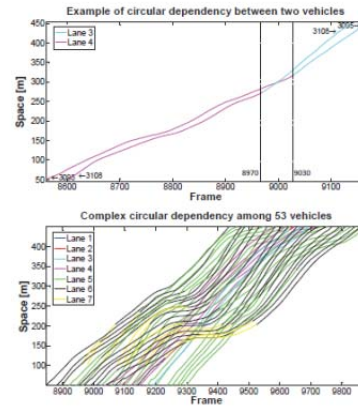
Removal of: 1) outliers and 2) noise



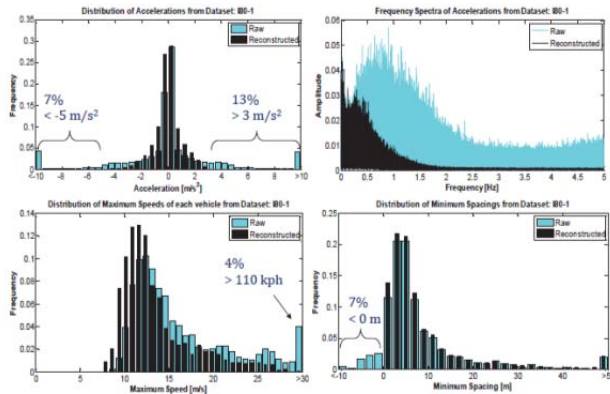
3) reconstruction 4) residual noise removal



Imposing platoon consistency

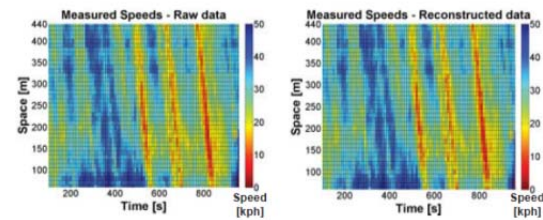


Raw vs. Reconstructed trajectories (I80-1)



Raw vs. reconstructed macroscopic patterns

resolution 10m x 10s



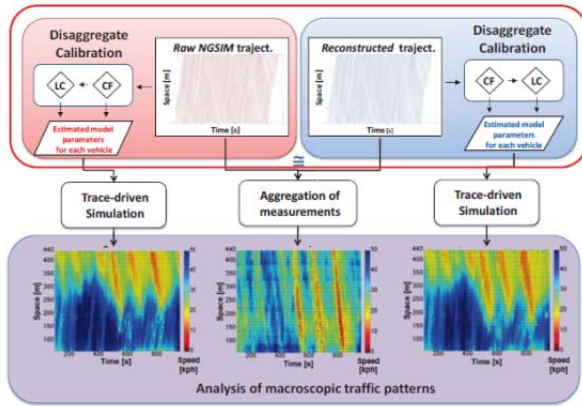
Aggregating trajectory data, differences are barely visible

Impacts of trajectory measurement errors: background

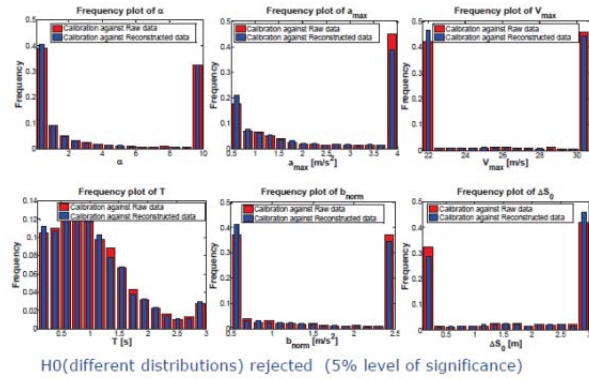
- Ossens and Hoogendoorn, 2008, TRR
 - ✓ Impact of errors on calibration of car-following models
 - ✓ Experiment with synthetic data
 - ✓ white Gaussian noise assumption
- Significant impact of errors on calibration results
- Only impact on calibration: no lane changing and no impact on aggregate results that is on traffic simulation outputs

IMPACT OF MEASUREMENT ERRORS ON CALIBRATION OF DRIVER MODELS (CF AND LC)

Impact of measurement errors on simulation

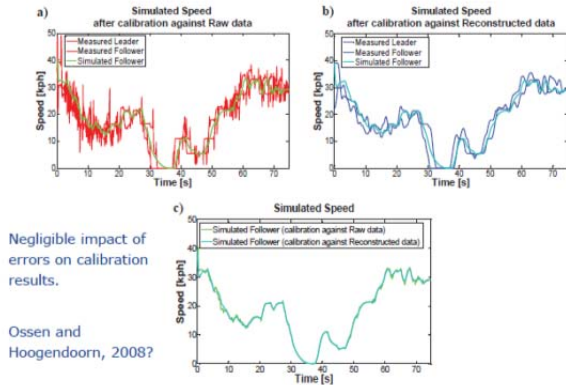


Frequencies of calibrated IDM parameters



H0(different distributions) rejected (5% level of significance)

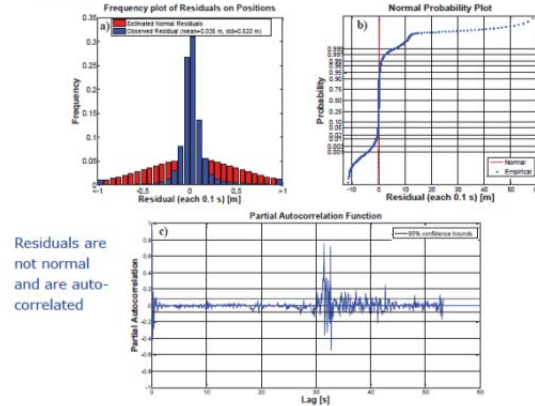
The car-following model acts as a filter



Negligible impact of errors on calibration results.

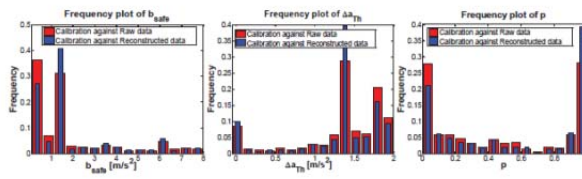
Ossen and Hoogendoorn, 2008?

Hp: Measurement errors = residuals between raw and reconstructed positions



Residuals are not normal and are auto-correlated

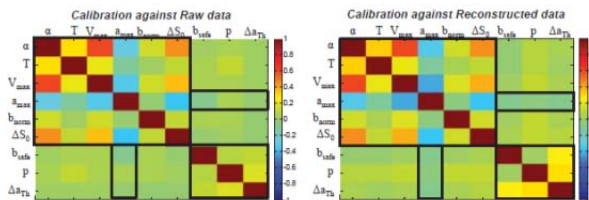
Frequencies of calibrated MOBIL parameters



H0(different distributions) not rejected (5% level of significance)

Correlation matrices of «raw and clean parameters» (i.e. parameters calibrated against raw and reconstructed data)

Analysis of parameters correlation (Kim and Mahmassani, 2012, TRR)



- Correlation among CF parameters does not change;
- correlation of a_{max} CF parameter with LC parameters increases (in abs.);
- correlation among LC parameters increases.
- Hp. explanatory capability of the LC model is reduced by errors

Summary of impacts on calibration

- Not normal, auto-correlated residuals
- No impact on car-following calibration results (both on parameters PDF and correlation structure)
- Impact on Lane-changing calibration results (both on parameters PDF and correlation structure)
- ...
- Analysis results not informative on the impacts on traffic simulation

Remarks

No car-following calibration possible on raw data without 'tricks' (negative inter-vehicle spacing)

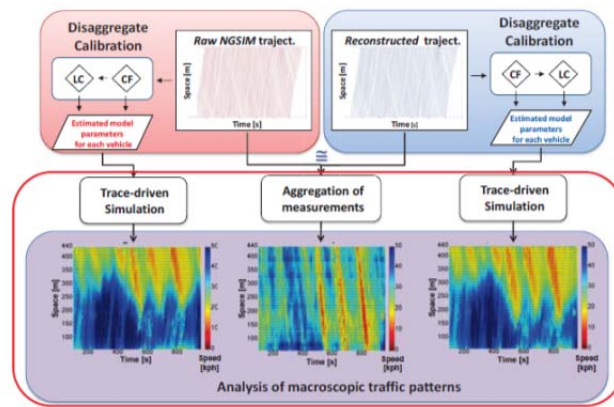
To calibrate a CF and LC model over a whole set of 2000 trajectories is not just running 2000 times an optimisation algorithm!

Major critical steps:

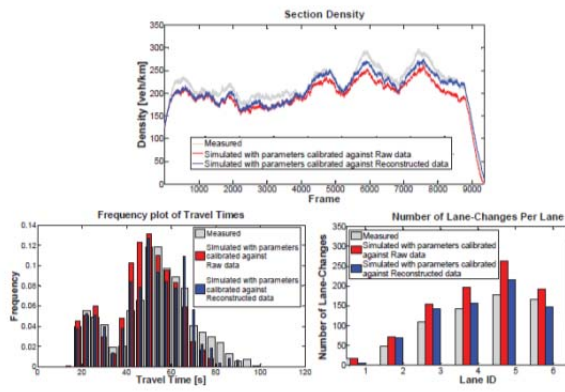
- To verify the calibration setting (MoP, GoF, algorithm)
- To set the bounds for the parameters (Punzo et al. 2014, IEEE T-ITS)
 - ✓ Uncertainty analysis
 - ✓ Physical criteria
 - ✓ Verification after simulation

IMPACT OF MEASUREMENT ERRORS ON SIMULATION OUTPUTS AND MACROSCOPIC TRAFFIC PATTERNS

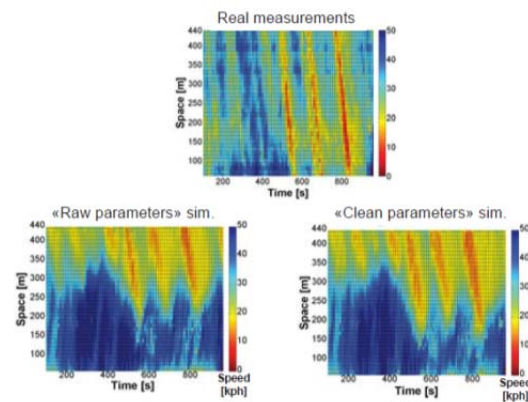
Impact of measurement errors on simulation



«Raw vs. clean parameters» simulations



Measured vs. simulated Edie's speed

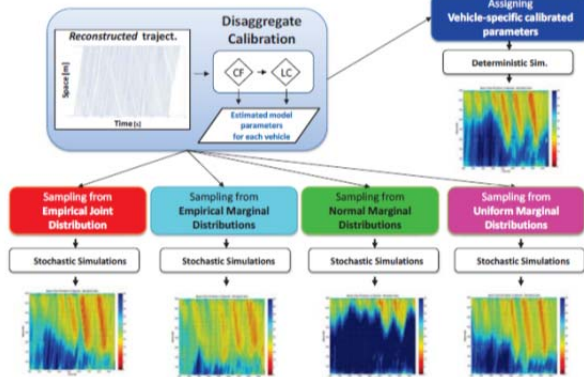


Remarks

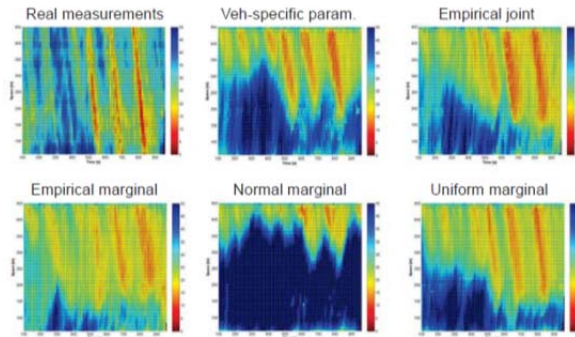
- Impacts of errors on simulation less than expected, given the big errors in raw data
- Measures other than macroscopic patterns could not capture the true behavior of models
- In both the simulations congestion patterns are different from that measured:
 - ✓ congestion is lower at the beginning of the stretch, higher at the end.
 - ✓ Models are not able to reproduce the full upstream propagation of congestion
- Yet, the 'clean data simulation' provided slightly better description of traffic at both disaggregate and collective description

IMPACTS OF ASSUMPTIONS ON THE INPUT PARAMETER PROBABILITY DENSITY FUNCTIONS (PDF)

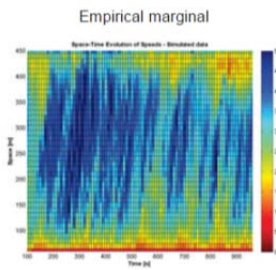
Assigning calibrated parameters to vehicles



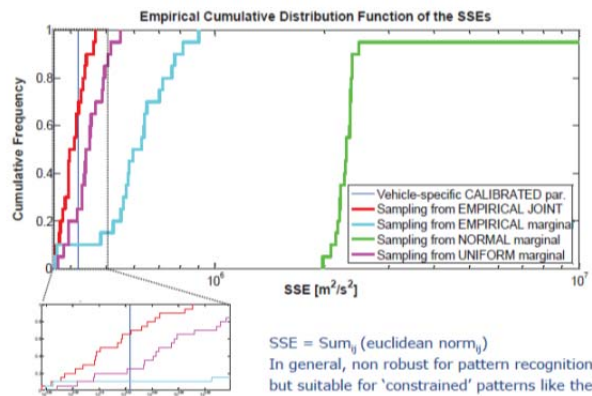
Real Measurements vs. best replication of scenarios



Virtual queue at the entrance



Empirical CDF of scenario replications' SSEs



Remarks

- Huge impact on results of the assumption on the input parameter PDFs
- CDFs on SSE offer a clear ranking of the performances of different input PDFs
- Sampling from 'empirical marginal' PDF (i.e. no correlation) yields 'unrealistic' parameters combinations (i.e. virtual queues at the entrance)
- Sampling from normal marginal, all the behaviours are averaged → no congestion propagation
- If the parameter correlation structure is unknown, the safest assumption is to sample the parameters from uniform PDFs (customary assumption in case of no prior information on the PDF)

Summary (2)

- Quantification of measurement errors impact on the collective behaviour of traffic;
- Analysis of the impact of assumptions on the probabilistic model of inputs on the simulation results, in terms of macroscopic spatiotemporal traffic patterns.

Main work references

Montanino M. and Punzo V. 2015. Errors in vehicle trajectory data and their impact on simulated spatiotemporal traffic patterns. Submitted for presentation to the 2015 TRB Annual meeting, Washington D.C.

Punzo V., M. Montanino and Ciuffo B., 2014. Do we really need to calibrate all the parameters? Variance-based sensitivity analysis to simplify microscopic traffic flow models. *IEEE Transactions on Intelligent Transportation Systems* (in press)

M. Montanino, V. Punzo, 2013. Making NGSIM data usable for studies on traffic flow theory: a multistep method for vehicle trajectory reconstruction. *Transportation Research Record* 2390, p.99-111, TRB.

Punzo V., Ciuffo B. and Montanino M. 2012. Can results of car-following model calibration based on trajectory data be trusted? *Transportation Research Record* 2315, p. 11-24. "Greenfields prize 2012"

Punzo V., Borzacchiello M., Ciuffo B., 2011. On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. *Transportation Research Part C* 19 (2011), 1243-1262.

www.multitude-project.eu

Summary (1)

- Impact of measurement errors in trajectory data substantially neglected in the field literature.
- Enhanced methodology to reconstruct trajectory data, accounting for inter-vehicle spacing consistency;
- Application to the NGSIM I80-1 dataset;
- Straightforward methodology to evaluate the impacts of microscopic features on collective behaviour of traffic based on:
 - ✓ consistent calibration of individual driver models (lane changing and car-following) over the entire trajectory dataset
 - ✓ trace-driven traffic simulation over the same time-space domain of trajectory data.
- ✓ New methodology to calibrate rule-based lane changing models (e.g. MOBIL), based on the concept of lane-changing scenario.
- ✓ Quantification of measurement errors impact on both car-following and lane changing calibration:
 - ✓ Negligible impact on CF
 - ✓ Quantified impact on LC

Conclusions

- Previous studies using NGSIM-like data rarely made use of collective description of traffic, as resulting from macroscopic spatiotemporal traffic patterns, to corroborate models and model assumptions
- NGSIM-like data open up new horizons in researching traffic flow theory and simulation, enabling the study of the collective behavior of traffic resulting from single driver models (i.e. car-following and lane-changing)
- Investigations made on NGSIM-like data will hopefully contribute to solve the dichotomy between TFT and traffic simulations as well as micro/macro dualism
- New data gathering efforts of NGSIM-like data are needed around the world

Contacts:

Vincenzo Punzo

University of Naples Federico II, Italy

vinpunzo@unina.it

KEYNOTE ADDRESS

Traffic Theory in the Era of Autonomous Vehicles

MICHAEL ZHANG
University of California, Davis

Traffic Flow Theory in the Era of Autonomous Vehicles

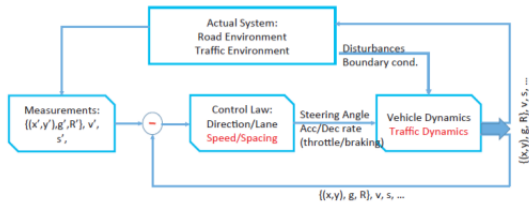
Michael Zhang
 University of California Davis

A Presentation at
 The Symposium Celebrating 50 Years of Traffic Flow Theory,
 August 11-13, 2014, Portland, Oregon USA

Outline

- From individual driving to traffic flow
- Prominent features of traffic flow
- Models of traffic flow-human driven vehicles
- Models of traffic flow-autonomous vehicles
- The future of traffic flow

The Driving Task as Feedback Control



Human Drivers vs Autonomous Vehicles From A Control Perspective

- | Human Drivers | Autonomous Vehicles (Robo Cars) |
|---|--|
| <ul style="list-style-type: none"> • Sensing is imprecise but more versatile • Response is slower but more robust • Best at processing fuzzy information and is highly adaptive • Strength: handles complex tasks such as lane tracking, obstacle avoidance more easily | <ul style="list-style-type: none"> • Sensing is more precise but less versatile • Response is faster but less robust • Best at exercising precise controls and is less adaptive: • Strength: handles procedural tasks such as speed control, car following more easily |

The Essence of Traffic Flow Theory is to Infer

- The Speed-Spacing Control Law of Each Driver

$$v_{\Delta n}(t) = \{?\}(s_{\Delta n}(t), \dots, E)$$

$E = \{\text{speed limits, grades, radius, surface conditions, visibility, } \dots\}$

- And the collective dynamics of an OPEN "Many-Particle" Dynamical System with "random" insertions and removals (reflecting LANE CHANGE interactions) controlled by these driver control laws

$$\{x_{\Delta n}(t) = v_{\Delta n}(t), n=1,2,\dots,N\}$$

Example: The California Motor Code Rule

- For every 10 mph of speed, leave one car length of space

- This translates to

$$s(t) - l = v(t)/10 \quad l \equiv T v(t)$$

or

$$v(t) = s(t) - l/T$$

with speed limits

$$v(t) = \min\{V_{\Delta f}, s(t) - l/T\}$$

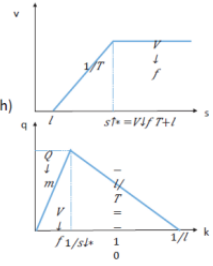
If Human Drivers are Identical Robots

with super fast reaction time and vehicles capable of infinite acceleration and deceleration

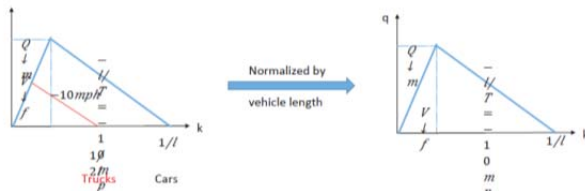
- Micro model $x(t)=v(t)T, v(t)=a(t)T, v(t)=Vlf @ u(t)-v(t)/T, v(t)=a(t)$
- Traffic Stream Model (steady-state) $V(s)=min[Vlf, s/T]$
- Macro (continuum) model (in vehicle coordinate) $s \downarrow t - v \downarrow n = 0, v=V(s)$

What are These Models and what phenomena do they produce?

- Micro model: "linear" CF model of Pipes
 - Acceleration waves
 - Deceleration waves
- Stream model: Triangular FD
 - Capacity: 2640 pcphpl ($l=20ft, T=1.36sec, Vf=60mph$)
 - Jam wave speed: -10 mph
- Macro model: LWR with Triangular FD
 - $klx + Qlx(k)=0$
 - Shock waves
 - Expansion (acceleration) waves



When All Vehicles Follow the Same Rule



The slope of the jam wave speed is a good indicator whether drivers of different type of vehicles follow the same driving rule or not

In reality, human drivers

- Differ from each other in driving ability and habits
- Cannot assess motion and distances precisely
- Respond with delay and finite acceleration/deceleration
- Do not follow rules exactly

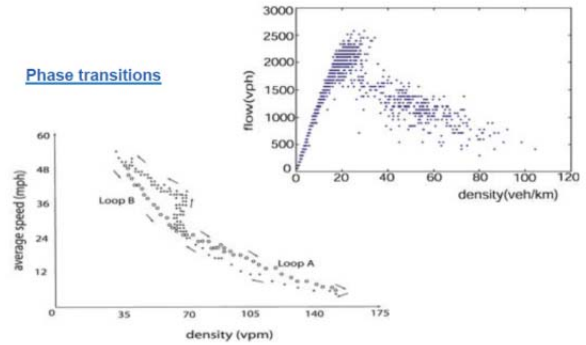
Consequence:

Traffic flow in the real world is much more complex

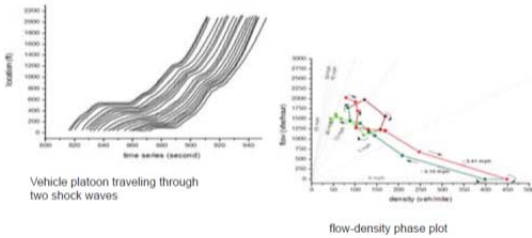
Prominent Features of Real Traffic Flow

- Phase transitions
- Nonlinear waves
- Stop-and-Go Waves (periodic motion)

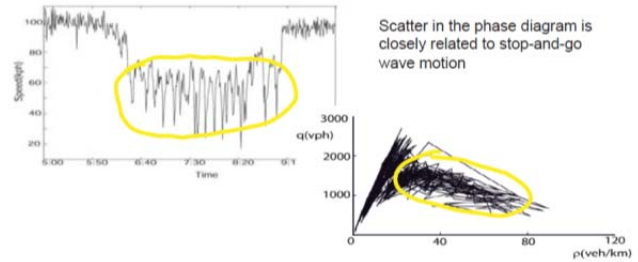
Phase transitions



Nonlinear waves



Stop-and-Go Waves (Oscillations)



Some Classical Traffic Models

- Microscopic
 - Modified Pipes' model
 - Newell' Model
 - Bando' model

$$\begin{aligned} \xi_i &= \min\{v_f, (s_n(t) - l)/\tau\} \\ \xi_i(t + \tau) &= v_f [1 - \exp\{-\lambda(s_n(t) - l)/v_f\}] \\ \xi_i(t) &= a [(u_n(s_n) - \xi_i)(t)], a = 1/\tau \end{aligned}$$

- Macroscopic continuum
 - LWR model
 - Payne-Whitham model
 - Aw-Rasclé, Zhang model

$$\rho = 1/s, u_n(s) = v_n(\rho), q = \rho v, q_n(\rho) = \rho v_n(\rho)$$

$$\begin{aligned} \rho_t + q_n(\rho)_x &= 0 \\ \rho_t + (\rho v)_x &= 0, \quad v_x + (v v)_x + \frac{c_0^2}{\rho} \rho_x = \frac{v_x(\rho) - v}{\tau} \\ \rho_t + (\rho v)_x &= 0, \quad v_x + (v - c(\rho))v_x = \frac{v_x(\rho) - v}{\tau} \\ c(\rho) &= -\rho v'_n(\rho) \end{aligned}$$

- v-s (speed-spacing) relation is central to all these models

The Difficulty of Modeling Real Flow

- Each driver is different
- Driving rules are hidden
- Sensing is imprecise
- Behavior is adaptive, nonlinear, and perhaps inconsistent
- (Driving environment is complex)

When Robo Cars Take Over the Road

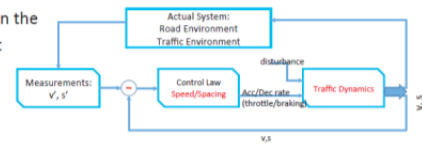
- Behavior is uniform and consistent
- Sensing and control is more precise
- Rules are always obeyed
- (Driving environment is still complex)

More importantly, driving rules are by design, leaving rooms for optimizing flow and safety → Feedback Control Problem

Traffic Flow Theory For Robo Cars- Longitudinal Control

- Example RoboCar#1
- $a(t + \tau) = k \downarrow r \{V(s) - v\}, V(s) = \min\{V \downarrow f, s/T\}$
- Human: $r=1-2s, T=1.36-2s$; Robo Car: $r=0.4-0.6s, T=0.8-1.2s$, Capacity: $\approx 1/T', +70\%$,

But this may be too rosy a prediction in the initial deployment stage (liability)



Example RoboCar#2

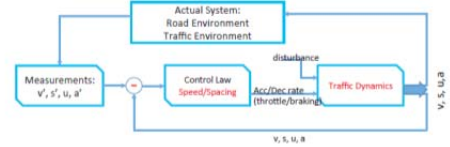
$$a(t + \tau) = k \downarrow r \{V(s) - v\} + k \downarrow v \{u - v\}$$

Faster response and higher throughput than RoboCar#1
 $r=0.4-0.6s, T=0.6-0.75s$



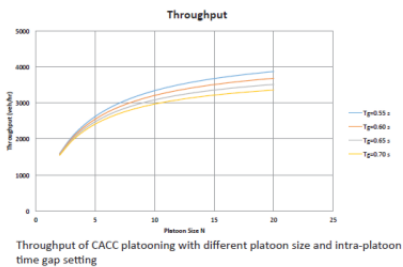
Example RoboCar#3 (RoboCar#2 with V2V)

$$a(t + \tau) = k \downarrow a \downarrow u(t) + k \downarrow r \{V(s) - v\} + k \downarrow v \{u - v\}$$



And the list goes on: you can come up with other models that meet safety and stability requirements

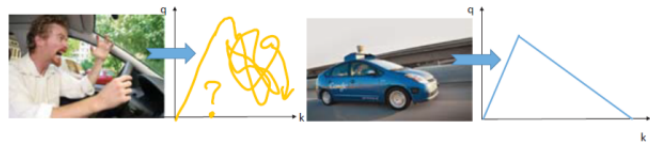
Expected throughput with vehicle platooning



Throughput of CACC platooning with different platoon size and intra-platoon time gap setting

Future of Traffic Flow Theory Research (1)

- Do the Arrival of Robo Cars Mean The End of Traffic Flow Research?
 - Automation creates uniformity and standardization, suppresses randomness: From billions of drivers to a handful: Google Car, GM Car, Toyota Car ...
 - Behavior of each Robo Car is consistent and known



From Human Drivers to Robots

Future of Traffic Flow Theory Research (2)

- In the short term
 - design of driving models for Robo cars
 - Robo car friendly infrastructure
- In the intermediate term
 - Mixed traffic with Robo Cars,
 - Platooning of Robo Cars
 - Lightless intersections with in-vehicle signal control
 - Rich micro level data for understanding and modeling traffic, and validating traffic models

Future of Traffic Flow Theory Research (3)

- In the long term, full automation of highway traffic
 - Optimal scheduling and pricing for congestion free networks
 - Robust Recovery from Disruptions
- New services and shared use of autonomous vehicles
 - Robo Taxi Services
 - Last and first-mile of transit (flexible transit)
 - Seamless integration of multiple modes
 - And the list goes on

Concluding Remarks

Autonomous Vehicles will

- In the long run bring more order to traffic flow and simplify traffic flow theory
- Produce rich data for traffic flow research
- Brings a host of brand new research problems for modeling, design and operations of transportation systems


KEYNOTE ADDRESS

Challenges in Pedestrian Flow Modeling
Macroscopic Modeling of Crowds Capturing Self-Organization

SERGE P. HOOGENDOORN
Delft University of Technology, the Netherlands



Challenges in Pedestrian Flow Modelling
 Macroscopic Modeling of Crowds capturing Self-Organisation
Serge Hoogendoorn, Winnie Daamen, Femke van Wageningen-Kessels, and others...





Challenges in Pedestrian Flow Modelling
 Macroscopic Modeling of Crowds capturing Self-Organisation
Serge Hoogendoorn, Winnie Daamen, Femke van Wageningen-Kessels, and others...

Societal Urgency
 Examples showing increasing importance of crowd modeling & management
Religious or social gatherings, events, (re-)urbanisation, increase use of transit and rail...

- Situations where large crowds gather are frequent (sports events, religious events, festivals, etc.); occurrence of 'spontaneous events' due to social media
- Transit (in particular train) is becoming more important leading to overcrowding of train stations under normal and exceptional situations (e.g. renovation of Utrecht Central Station)
- Walking (and cycling) are becoming more important due to (re-)urbanisation (strong reduction car-mobility in Dutch cities)
- When things go wrong, societal impacts are huge!



Courtesy of Prof. H. Mahmassani

How can models be used to support planning, organisation, design, and control?

- Testing (new) designs of stations, buildings, stadiums, etc.
- Testing evacuation plans for buildings
- Testing crowd management and control measures and strategies
- Training of crowd managers
- On-line crowd management systems as part of state estimation and prediction

Development of valid models requires good data!!!

Engineering challenges
 Societal urgencies leads to demand for engineering solutions
Importance of Theory and Models for Pedestrian and Crowd Dynamics



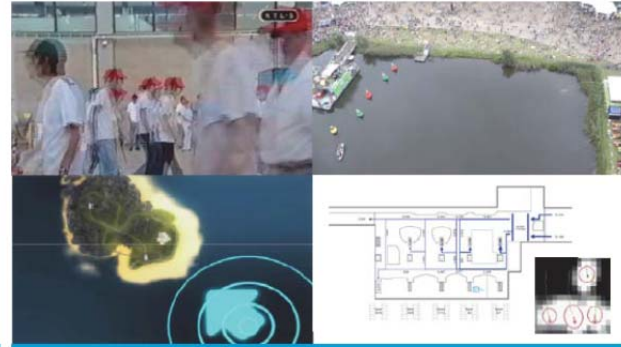
How can models be used to support planning, organisation, design, and control?

- Testing (new) designs of stations, buildings, stadiums, etc.
- Testing evacuation plans for buildings
- Testing crowd management and control measures and strategies
- Training of crowd managers
- On-line crowd management systems as part of state estimation and prediction

Development of valid models requires good data!!!

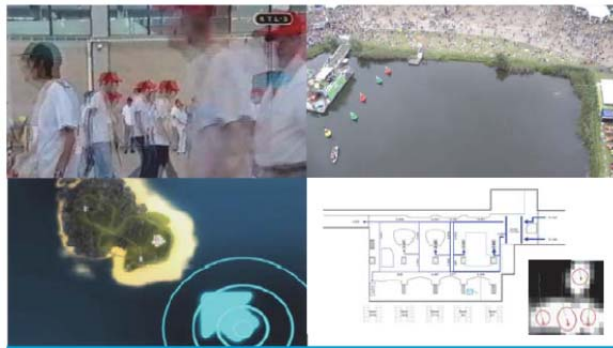
Engineering challenges

Societal urgencies leads to demand for engineering solutions
 Importance of Theory and Models for Pedestrian and Crowd Dynamics



Understanding Pedestrian Flows

Field observations, controlled experiments, virtual laboratories
 Data collection remains a challenge, but many new opportunities arise!

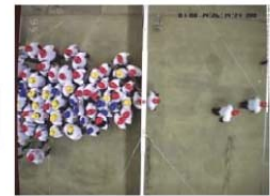


Understanding Pedestrian Flows

Field observations, controlled experiments, virtual laboratories
 Data collection remains a challenge, but many new opportunities arise!

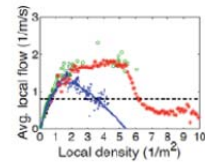
Empirical characteristics and relations

- Experimental research capacity values:
- $$C = 2.69 + 1.06 \cdot P_c - 0.21 \cdot P_s - 2.13 \cdot P_o - 0.01 \cdot \text{Stress} - 0.12 \cdot \text{Width} - 0.18 \cdot \text{Door} + 0.09 \cdot \text{Light}$$
- Strong influence of composition of flow
 - Importance of geometric factors



Fundamental diagram pedestrian flows

- Relation between density and flow / speed
- Big influence of context!
- Example shows regular FD and FD determined from Jamarat Bridge

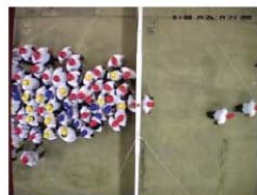


Traffic flow characteristics for pedestrians...

Capacity, fundamental diagram, and influence of context

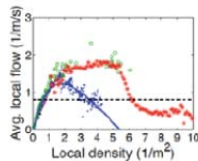
Empirical characteristics and relations

- Experimental research capacity values:
- $$C = 2.69 + 1.06 \cdot P_c - 0.21 \cdot P_s - 2.13 \cdot P_o - 0.01 \cdot \text{Stress} - 0.12 \cdot \text{Width} - 0.18 \cdot \text{Door} + 0.09 \cdot \text{Light}$$
- Strong influence of composition of flow
 - Importance of geometric factors



Fundamental diagram pedestrian flows

- Relation between density and flow / speed
- Big influence of context!
- Example shows regular FD and FD determined from Jamarat Bridge



Traffic flow characteristics for pedestrians...

Capacity, fundamental diagram, and influence of context

Phenomena in pedestrian flow operations

Fascinating world of pedestrian flow dynamics!



Examples self-organisation


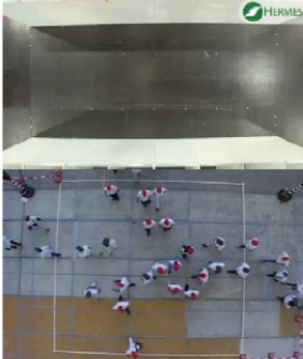
- Self-organisation of dynamic lanes in bi-directional flow
- Formation of diagonal stripes in crossing flows
- Viscous fingering in multi-directional flows

Characteristics:

- Self-organisation yields moderate reduction of flow efficiency
- Chaotic features, e.g. multiple 'stable' patterns may result
- Limits of self-organisation

Phenomena in pedestrian flow operations

Fascinating world of pedestrian flow dynamics!

Examples self-organisation


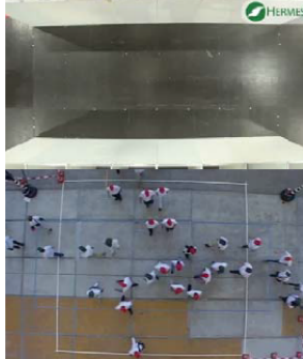
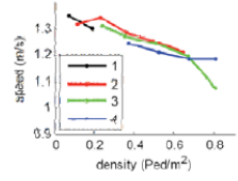
- Self-organisation of dynamic lanes in bi-directional flow
- Formation of diagonal stripes in crossing flows
- Viscous fingering in multi-directional flows

Characteristics:

- Self-organisation yields moderate reduction of flow efficiency
- Chaotic features, e.g. multiple 'stable' patterns may result
- Limits of self-organisation

Phenomena in pedestrian flow operations

Fascinating world of pedestrian flow dynamics!







Characteristics:

- Self-organisation yields moderate reduction of flow efficiency
- Chaotic features, e.g. multiple 'stable' patterns may result
- Limits of self-organisation

Limits to efficient self-organisation

Overloading causes phase transitions

Examples failing self-organisation

- When conditions become too crowded efficient self-organisation 'breaks down'
- Flow performance (effective capacity) decreases substantially, causing cascade effect as demand stays at same level
- New phases make occur: start-stop waves, **turbulence**

Network level characteristics

- Generalised network fundamental diagram can be sensibly defined for pedestrian flows!

Limits to efficient self-organisation

Overloading causes phase transitions




Examples failing self-organisation

- When conditions become too crowded efficient self-organisation 'breaks down'
- Flow performance (effective capacity) decreases substantially, causing cascade effect as demand stays at same level
- New phases make occur: start-stop waves, **turbulence**

Network level characteristics

- Generalised network fundamental diagram can be sensibly defined for pedestrian flows!

The Modelling Challenge

Reproducing key phenomena in pedestrian dynamics
Towards useful pedestrian flow models...

Challenge is to come up with a model that can reproduce or predict pedestrian flow dynamics under a variety of circumstances and conditions

Inductive approach: when designing a model, consider the following:

- Which are the key phenomena / characteristics you need to represent?
- Which theories could be used to represent these phenomena?
- Which mathematical constructs are applicable and useful?
- Which representation levels are appropriate
- How to tackle calibration and validation?



The Modelling Challenge

Reproducing key phenomena in pedestrian dynamics
Towards useful pedestrian flow models...

Challenge is to come up with a model that can reproduce or predict pedestrian flow dynamics under a variety of circumstances and conditions

Inductive approach: when designing a model, consider the following:

- Which are the key phenomena / characteristics you need to represent?
- Which theories could be used to represent these phenomena?
- Which mathematical constructs are applicable and useful?
- Which representation levels are appropriate
- How to tackle calibration and validation?



Pedestrian modelling approaches

Representation and behaviour roles

Examples of micro, meso, and macroscopic pedestrian modelling approaches

Representation:	Individual particles	Continuum
Behavioural rules:		
Individual behaviour (predictive)	Microscopic simulation models (social forces, CA, NOMAD, etc.)	Gas-kinetic models (Hoogendoorn, Helbing)
Aggregate behaviour (descriptive)	Particle discretisation models (SimPed)	Macroscopic continuum models (Hughes, Colombo, Hoogendoorn)

Research emphasis on microscopic simulation models and on walking behaviour

Example: NOMAD Game Theoretical Model

Interaction modelling by using differential game theory

Or: Pedestrian Economicus as main theoretical assumption...

Application of differential game theory:

- Pedestrians minimise predicted walking cost, due to straying from intended path, being too close to others / obstacles and effort, yielding $a(t)$:

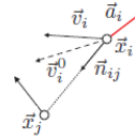
$$\vec{a}_i = \frac{\vec{v}_i^0 - \vec{v}_i}{\tau_i} - A_i \sum_j \exp\left[-\frac{R_{ij}}{B_i}\right] \cdot \vec{n}_{ij} \cdot \left(\lambda_i + (1 - \lambda_i) \frac{1 + \cos \phi_{ij}}{2}\right)$$

Level of anisotropy reflected by this parameter

- Simplified model is similar to Social Forces model of Helbing

Face validity?

- Model results in reasonable fundamental diagrams
- What about self-organisation?

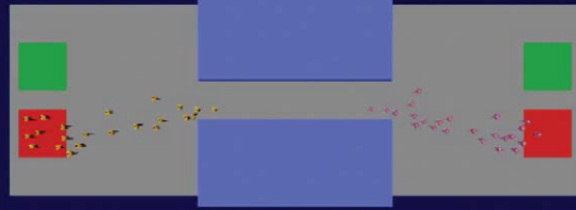


Example: NOMAD Game Theoretical Model

Interaction modelling by using differential game theory

Or: Pedestrian Economicus as main theoretical assumption...

Example shows lane formation process for homogeneous groups...



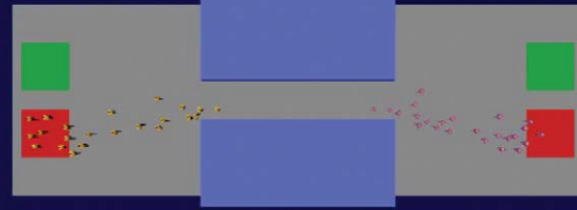
Heterogeneity yields less efficient lane formation (freezing by heating)

Example: NOMAD Game Theoretical Model

Interaction modelling by using differential game theory

Or: Pedestrian Economicus as main theoretical assumption...

Example shows lane formation process for homogeneous groups...



Heterogeneity yields less efficient lane formation (freezing by heating)

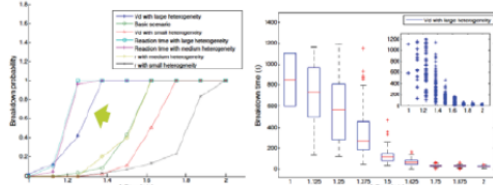
Example: NOMAD Game Theoretical Model

Interaction modelling by using differential game theory

Or: Pedestrian Economicus as main theoretical assumption...

Is NOMAD able to reproduce breakdown in case of oversaturation?

- Testing NOMAD / Social Forces model using different demand patterns to investigate if and under which conditions breakdown occurs
- Large impact of population heterogeneity ('freezing by heating') + reaction time



¹⁾ Work by Xiaodan Yang, Wim de Dreuven, Serge Paul Hoogendoorn, Yao Chen, Haihong Dong

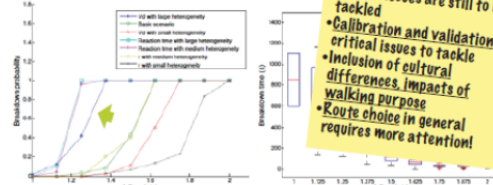
Example: NOMAD Game Theoretical Model

Interaction modelling by using differential game theory

Or: Pedestrian Economicus as main theoretical assumption...

Is NOMAD able to reproduce breakdown in case of oversaturation?

- Testing NOMAD / Social Forces model using different demand patterns to investigate if and under which conditions breakdown occurs
- Large impact of population heterogeneity ('freezing by heating')



Microscopic models have been relatively successful, although several issues are still to be tackled
 • Calibration and validation are critical issues to tackle
 • Inclusion of cultural differences, impacts of walking purpose
 • Route choice in general requires more attention!

¹⁾ Work by Xiaodan Yang, Wim de Dreuven, Serge Paul Hoogendoorn, Yao Chen, Haihong Dong

Dynamic programming

- NOMAD route choice: $\vec{v}_i^0 = \vec{\gamma}^0 \cdot V^0$

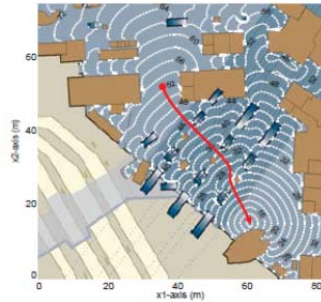
- Let $W(t, \mathbf{x})$ denote minimum expected cost (travel time) from (t, \mathbf{x}) to destination(s)

- $W(t, \mathbf{x})$ satisfies HJB equation:

$$\frac{\partial W}{\partial t} = L(t, \vec{x}, \vec{v}^0) + \vec{v}^0 \cdot \nabla W + \frac{\sigma^2}{2} \Delta W$$

where $\vec{\gamma}^0 = -\nabla W / \|\nabla W\|$

- Optimal direction $\vec{\gamma}^0$ perpendicular to iso-cost curves
- Efficient numerical solution approaches available...



Completing the Model

Route choice modelling by Stochastic Optimal Control

Optimal routing in continuous time and space...

Continuum modelling

Dynamic assignment in continuous time and space

Macroscopic traffic flow modelling...

Multi-class macroscopic model of Hoogendoorn and Bovy (2004)

- Kinematic wave model for pedestrian flow for each destination d

$$\frac{\partial \rho_d}{\partial t} + \nabla \cdot \vec{q}_d = r - s \quad \text{with} \quad \vec{q}_d = \vec{\gamma}_d \cdot \rho_d \cdot V(\rho_1, \dots, \rho_D)$$

- Here V is the (multi-class) equilibrium speed; the optimal direction:

$$\vec{\gamma}_d(t, \vec{x}) = -\frac{\nabla W_d(t, \vec{x})}{\|\nabla W_d(t, \vec{x})\|}$$

- Stems from minimum cost $W_d(t, \mathbf{x})$ for each (set of) destination(s) d
- Is this a reasonable model?

Continuum modelling

Dynamic assignment in continuous time and space

Macroscopic traffic flow modelling...

Multi-class macroscopic model of Hoogendoorn and Bovy (2004)

- Kinematic wave model for pedestrian flow for each destination d

$$\frac{\partial \rho_d}{\partial t} + \nabla \cdot \vec{q}_d = r - s \quad \text{with} \quad \vec{q}_d = \vec{\gamma}_d \cdot \rho_d \cdot V(\rho_1, \dots, \rho_D)$$

- Here V is the (multi-class) equilibrium speed; the optimal direction:

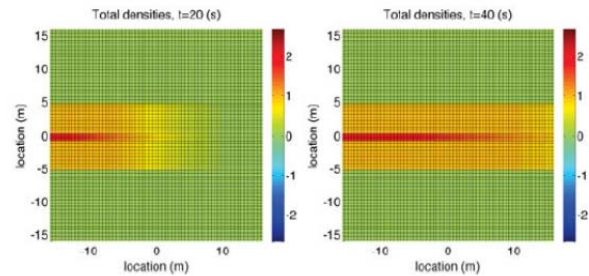
$$\vec{\gamma}_d(t, \vec{x}) = -\frac{\nabla W_d(t, \vec{x})}{\|\nabla W_d(t, \vec{x})\|}$$

- Stems from minimum cost $W_d(t, \mathbf{x})$ for each (set of) destination(s) d
- Is this a reasonable model?
- No, since there is only pre-determined (global) route choice, the model will have unrealistic features

Continuum modelling

Dynamic assignment in continuous time and space

Macroscopic traffic flow modelling...



Solution? Include a term describing local route / direction choice...

Continuum modelling - part 2

Computationally efficient modelling

Connecting microscopic to macroscopic models...

- NOMAD / Social-forces model as starting point:

$$\vec{a}_i = \frac{\vec{v}_i^0 - \vec{v}_i}{\tau_i} - A_i \sum_j \exp\left[-\frac{R_{ij}}{B_i}\right] \cdot \vec{n}_{ij} \cdot \left(\lambda_i + (1 - \lambda_i) \frac{1 + \cos \phi_{ij}}{2}\right)$$

- Equilibrium relation stemming from model ($\mathbf{a}_i = 0$):

$$\vec{v}_i = \vec{v}_i^0 - \tau_i A_i \sum_j \exp\left[-\frac{R_{ij}}{B_i}\right] \cdot \vec{n}_{ij} \cdot \left(\lambda_i + (1 - \lambda_i) \frac{1 + \cos \phi_{ij}}{2}\right)$$

- Interpret density as the 'probability' of a pedestrian being present, which gives a **macroscopic equilibrium relation** (expected velocity), which equals:

$$\vec{v} = \vec{v}^0(\vec{x}) - \tau A \iint_{\vec{y} \in \Omega(\vec{x})} \exp\left(-\frac{\|\vec{y} - \vec{x}\|}{B}\right) \left(\lambda + (1 - \lambda) \frac{1 + \cos \phi_{xy}(\vec{v})}{2}\right) \frac{\vec{y} - \vec{x}}{\|\vec{y} - \vec{x}\|} \rho(t, \vec{y}) d\vec{y}$$

- Combine with conservation of pedestrian equation yields complete model, but numerical integration is computationally very intensive

Continuum modelling - part 2

Computationally efficient modelling

Connecting microscopic to macroscopic models...

- Taylor series approximation:

$$\rho(t, \vec{y}) = \rho(t, \vec{x}) + (\vec{y} - \vec{x}) \cdot \nabla \rho(t, \vec{x}) + O(\|\vec{y} - \vec{x}\|^2)$$

yields a closed-form expression for the equilibrium velocity $\vec{v} = \vec{e} \cdot V$, which is given by the equilibrium speed and direction:

$$V = \|\vec{v}^0 - \beta_0 \cdot \nabla \rho\| - \alpha_0 \rho$$

$$\vec{e} = \frac{\vec{v}^0 - \beta_0 \cdot \nabla \rho}{V + \alpha_0 \rho}$$

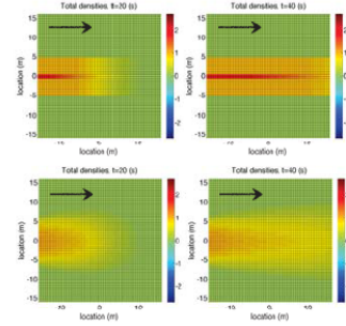
- with: $\alpha_0 = \pi \tau A B^2 (1 - \lambda)$ and $\beta_0 = 2 \pi \tau A B^3 (1 + \lambda)$
- Check behaviour of model by looking at isotropic flow ($\lambda = 1$) and homogeneous flow conditions ($\nabla \rho = \vec{0}$)
- Multi-class generalisation + Godunov scheme numerical approximation

Continuum modelling - part 2

Computationally efficient modelling

Connecting microscopic to macroscopic models...

- Uni-directional flow situation
- Picture shows differences between situation without and with local route choice for two time instances
- Model introduces 'lateral diffusion' since pedestrians will look for lower density areas actively
- Diffusion can be controlled by choosing parameters differently
- Model shows plausible behaviour

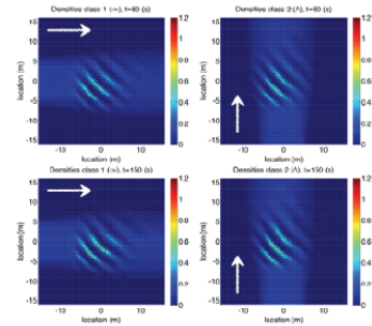


Continuum modelling - part 2

Computationally efficient modelling

Connecting microscopic to macroscopic models...

- Simulation results also show formation of diagonal stripes...
- Patterns which are formed depend on parameters of models
- In particular, non-equal impact of own class and other classes on diversion behaviour appears important

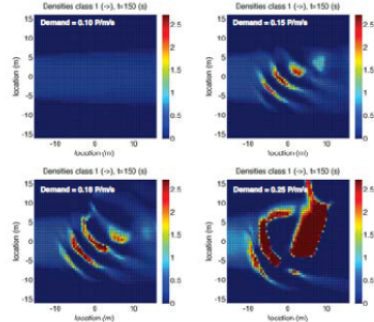


Continuum modelling - part 2

Computationally efficient modelling

Connecting microscopic to macroscopic models...

- Whether self-organisation occurs depends on demand level
- Low demand levels, no self-organisation
- Self-organisation fails for high demands and results in complete grid-lock (no outflow)
- Macroscopic model appears able to qualitatively reproduce crowd characteristics!

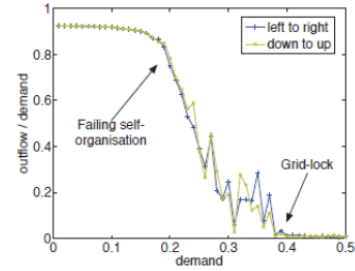


Continuum modelling - part 2

Computationally efficient modelling

Connecting microscopic to macroscopic models...

- Whether self-organisation occurs depends on demand level
- Low demand levels, no self-organisation
- Self-organisation fails for high demands and results in complete grid-lock (no outflow)
- Macroscopic model appears able to qualitatively reproduce crowd characteristics!

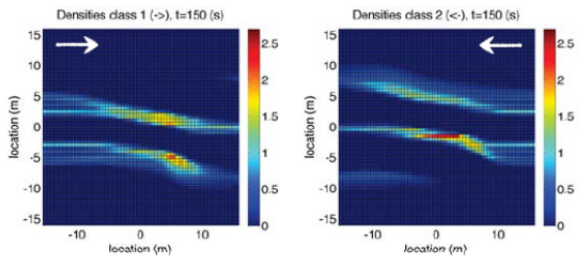


Continuum modelling - part 2

Computationally efficient modelling

Connecting microscopic to macroscopic models...

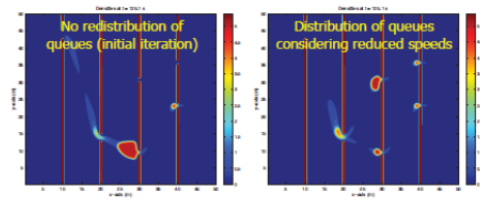
- Model seems to reproduce self-organised patterns (e.g. example below shows lane formation for bi-directional flows)



Applications?

Use of macroscopic flow model in optimisation

- Work presented at TRB 2013 proposes optimisation technique to minimise evacuation times
- Bi-level approach combining optimal routing (HJB equation) and continuum flow model (presented here)
- Preliminary results are very promising






KEYNOTE ADDRESS

Trajectories in 3-D

Unifying Model Calibration and Network Performance Analysis


HANI S. MAHMASSANI
Northwestern University



McCormick
 Northwestern Engineering
 Northwestern University Transportation Center


**TRAJECTORIES IN 3D:
 UNIFYING MODEL CALIBRATION AND NETWORK
 PERFORMANCE ANALYSIS**

Hani S. Mahmassani



Northwestern University

Symposium Celebrating 50 Years of Traffic Flow Theory
 Portland, Oregon (USA), August 11-13, 2014




McCormick
 Northwestern Engineering
 Northwestern University Transportation Center

Contributors:

Jiwon Kim *(now at U. of Queensland, Brisbane, AU)*
 Tian Hou
 Meead Saberi *(now at Monash Univ., Melbourne, AU)*
 Ali Zockaie

Doctoral researchers at NUTC



McCormick
 Northwestern Engineering
 Northwestern University Transportation Center

Outline

- Motivation
- From 2D trajectories to 3D trajectories
- Application to Network Flow Modeling (NFD's)
 - Vehicular networks
 - Pedestrians and crowds
- Travel time reliability
 - Signature relations and trajectory data
 - Within-day and day-to-day variability
- Scenario-based approach to reliability modeling
 - Trajectory Processor for particle-based simulators
- Takeways, Limitations and Challenges

Motivation

- NETWORK TRAFFIC FLOW MODELING needs high quality traffic data with broad network level coverage, for calibration, validation, and input to real-time predictive management strategies.
- CHARACTERIZATION OF NETWORK PERFORMANCE, and the quality of service experienced by users increasingly encompasses broader array of dimensions— e.g. reliability – that call for tracking vehicles as they travel through the network, and not only as they pass selected points
- Development of telecommunication and wireless technology are augmenting conventional point-based data collection methods with low-cost and widely available probe data.



Motivation II

- **LOOKING AHEAD—**
Autonomous Vehicles and Connected Vehicles/Systems will play a growing role part within the advanced traffic data environment, both as a major generator (data source) as well as end-user.



5

Segment Traffic Data

- **Segment Data**
 - collected by electronic transponders
 - Automatic Vehicle Identification (AVI), electronic toll data (I-PASS), blue tooth data, etc



Trajectory Data

- Collected by probe vehicles equipped with on board GPS devices
- A trajectory is the path followed by the moving object through the spatial area over which it moves



7

Trajectory Data

- **Information that can be extracted from trajectory data**
 - from individual trajectory:
 - Time, i.e. position of this moment on the timescale;
 - Position of the vehicle in space;
 - Trip origins and destinations ;
 - Direction of the vehicle's movement;
 - Speed of the movement;
 - Dynamics of the speed (acceleration/deceleration);
 - Accumulated travel time and distance.
 - Individual path and temporal characteristics
 - from groups of trajectories:
 - Distribution of speed/travel time;
 - Probe vehicle density;
 - Inferred traffic volume.

8

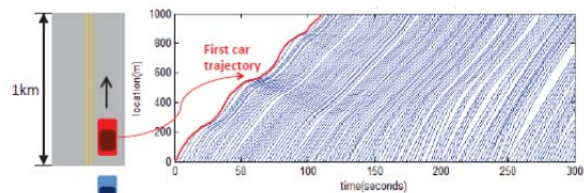
Trajectory Data

- Advantages and limitations of trajectory data as compared to traditional traffic data

Advantages	Limitations
<ul style="list-style-type: none"> • Low or no cost in installation and maintenance; • Wider geographic coverage (freeways and arterials); • Finer resolution (individual vehicle and shorter measurement time interval); • Contains additional traffic information (e.g. travel time); • Not affected by traffic interruptions or bad weather conditions. • Traffic simulation tools (microscopic, mesoscopic, or "particle-based" simulators) naturally produce trajectories 	<ul style="list-style-type: none"> • Technology is not as mature as fixed sensors; • No direct occupancy or traffic density information; • Limited experience in analyzing data.

9

2D Trajectories

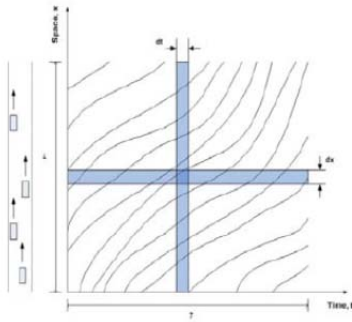


2D trajectories (along segment) have played essential role in development of traffic theories for individual highway facilities.

However, in validation and application of traffic simulation models, the focus has been on measurements taken at a point (using fixed sensors)

10

Measurements from Multiple Trajectories along a Single Road Segment



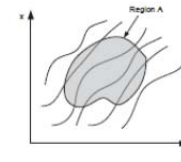
11

Edie's Definitions

Highway traffic (unidirectional)

Trajectory-based definitions of network flow variables

$$q(A) = \frac{\sum_{n \in N} d_n}{|A|} \quad k(A) = \frac{\sum_{n \in N} \tau_n}{|A|}$$



where d_n is the total distance traveled by vehicle n in region A , τ_n is the total time spent by vehicle n in region A , and $|A|$ is the area covered by region A .

Edie (1965)

12

Network Fundamental Diagram

Trajectory-based definitions

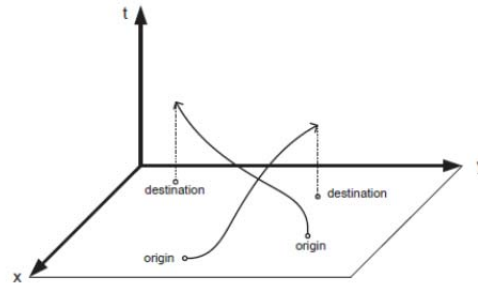
It is possible to extend Edie's well-known generalized variable definitions of vehicle traffic flow along a highway to a network, as recently recognized by Courbon and Leclercq (2011).

Recently, in Saberi, Mahmassani and Zockaie (2014)

- Operationalize and validate the extension of Edie's definitions to the network level.
- Formalize and test a method using three-dimensional (3D) vehicle trajectories in time and space to estimate network flow, density, and speed.

13

3D Trajectories in a Network

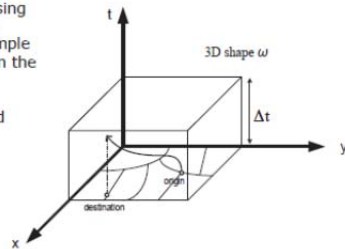


14

Network 3D Time-Space Diagram

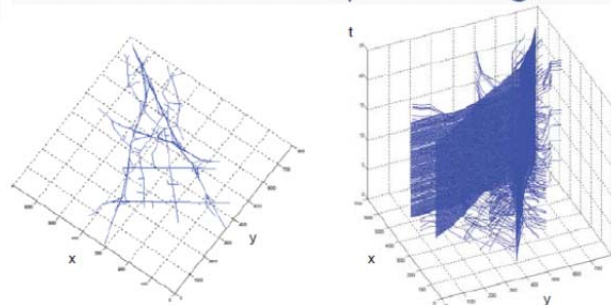
In order to estimate network-wide traffic flow variables using trajectories, we introduce a closed 3D shape ω , for example a cube, similar to region A in the 2D time-space diagram.

The network structure is laid down on the x - y plane.



15

Network 3D Time-Space Diagram



3D trajectories of 1,000 simulated vehicles in Irvine, California

Saberi, Mahmassani, Zockaie (2014)

16

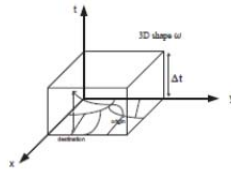
Edie's Definitions

Extension to Networks

*Courbon and Leclercq (2011)
Saberi, Mahmassani, Zockaie (2014)*

$$Q(\omega) = \frac{d(\omega)}{L_{xy}(\omega) \times \Delta t}$$

$$K(\omega) = \frac{t(\omega)}{L_{xy}(\omega) \times \Delta t}$$



where $Q(\omega)$ and $K(\omega)$ are the network-wide average flow and density for the specified shape ω ; $d(\omega)$ is the total distance traveled by all the vehicles in the shape ω , $t(\omega)$ is the total time spent by all vehicles in the shape ω , $L_{xy}(\omega)$ is the total length (in lane-miles or lane-kms) of the network on the x-y plane associated with the shape ω , and Δt is the time height of the shape ω .

17

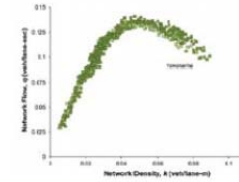
Background

Network Fundamental Diagram

Link-based definitions

Most of the studies to date have used the classical link-based measurement method to estimate the NFD by taking the distance-weighted averages of flow and density over all the links in the network.

$$Q = \frac{\left(\sum_{i=1}^M l_i q_i\right)}{\left(\sum_{i=1}^M l_i\right)} \quad K = \frac{\left(\sum_{i=1}^M l_i k_i\right)}{\left(\sum_{i=1}^M l_i\right)}$$



Source: Geroliminis and Daganzo (2008)

Source: Mahmassani et al. (1984)

18

Network Fundamental Diagram

Trajectory-based definitions

It is possible to extend Edie's well-known generalized variable definitions of vehicle traffic flow along a highway to a network, as recently recognized by Courbon and Leclercq (2011):

$$\bar{q}(t \rightarrow t + \Delta t, x \rightarrow x + \Delta x) = \frac{\sum_i l_i}{\Delta t \Delta x}$$

$$\bar{k}(t \rightarrow t + \Delta t, x \rightarrow x + \Delta x) = \frac{\sum_i t_i}{\Delta t \Delta x}$$

where l_i and t_i are respectively the distance traveled and the time spent by vehicle i in a time-space area of $\Delta x, \Delta t$.

19

Network Traffic Simulation

Networks of Chicago and Salt Lake City are simulated in a simulation-based dynamic traffic assignment platform (DYNASMART-P) with normal daily demand and 20% adaptive drivers to prevent formation of large gridlock.



Smaller Chicago sub-network



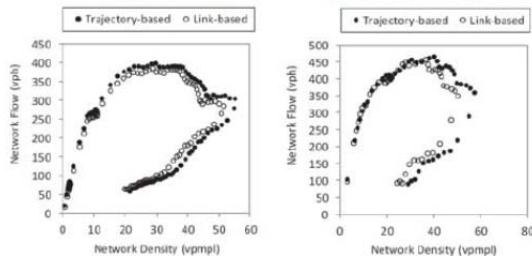
Salt Lake City sub-network

20

Trajectory vs. Link based NFD

Chicago Network

Salt Lake City Network



NFDs with Different Measurement Methods

21

Trajectory vs. Link based NFD

- In both networks, for network densities greater than 20 vpmpl, the link-based method underestimates average network densities.
- Both estimation methods yield near-identical network flows.
- When densities are high, the link-based method does not fully capture the variability of the congestion effects in the network.
- Averaging the number of simulated vehicles on individual links in each time interval creates a bias in the link-based method when estimating network densities.

22

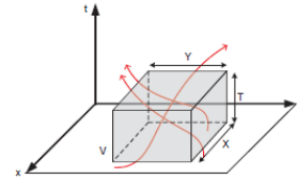


Outline

- > Motivation
- > From 2D trajectories to 3D trajectories
- > Application to Network Flow Modeling (NFD's)
 - Vehicular networks
 - Pedestrians and crowds
- > Travel time reliability
 - Signature relations and trajectory data
 - Within-day and day-to-day variability
- > Scenario-based approach to reliability modeling
 - Trajectory Processor for particle-based simulators
- > Takeaways, Limitations and Challenges

23

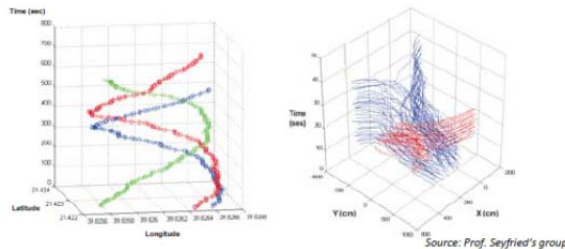
3D Time-Space Diagram



The three-dimensional time-space diagram of walking areas can be defined as a space in which the x and y axes represent the walking surface and the z axis represents time.

24

3D Time-Space Diagram



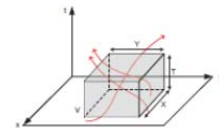
Three-Dimensional Illustration of (left) Pilgrims Trajectories in a Circular Environment (n=3) and (right) Pedestrian Trajectories in a Bidirectional Environment (n=100)

25

Edie's Definitions

Extension to multi-directional pedestrian areas

$$k = \frac{\sum_{n \in N} \tau_n}{|V|} = \frac{\sum_{n \in N} \tau_n}{T \cdot |A|} = \frac{\sum_{n \in N} \tau_n}{T \cdot (X \cdot Y)}$$



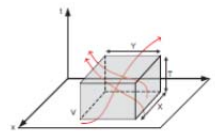
where τ_n is the total time spent by pedestrian n in shape V and $|V|$ is the spatial volume covered by shape V . Also, $|V|$ can be expressed as the geometric area of the walking area ($|A|=X \cdot Y$) multiplied by the time interval $T=(t_1-t_0)$.

26

Edie's Definitions

Extension to multi-directional pedestrian areas

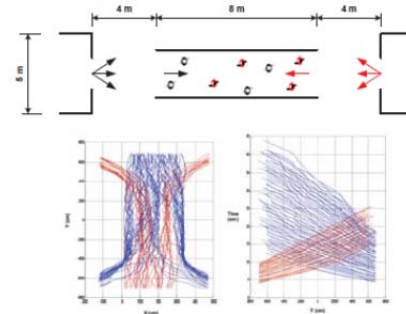
$$q = \frac{\sum_{n \in N} d_n}{|V|} = \frac{\sum_{n \in N} d_n}{Y \cdot X \cdot T}$$



where d_n is the total distance traveled by pedestrian n in shape V and $|V|$ is the spatial volume covered by shape V .

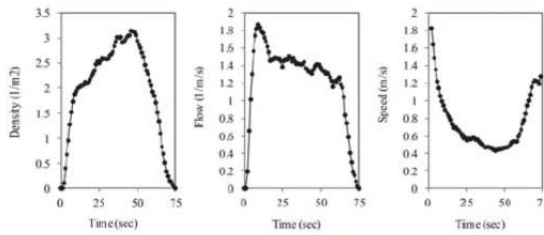
27

Experimental Data



Zhang et al. (2011)

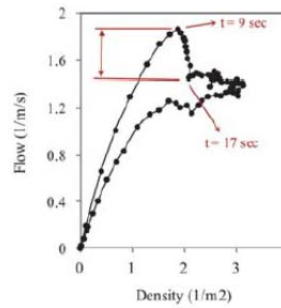
Pedestrian Traffic Measures



Time series of area-wide density, flow, and speed

29

Area-wide Fundamental Diagram



Similar to vehicular traffic flow on both individual facilities and networks, pedestrian traffic exhibits **hysteretic behavior** too.

The **capacity drop** phenomenon seems to exist in pedestrian crowds too.

The observed capacity drop ($t = 9-17$ sec) is followed by a relatively **stable period** in which the area-wide flow remains roughly constant while density continues to increase due to **formation of stable self-organized lanes**.

30

M^cCoormioik
Northwestern Engineering
Northwestern University Transportation Center

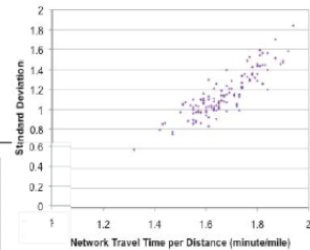
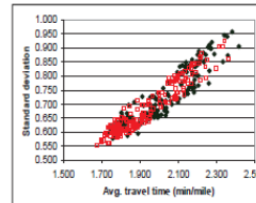
Outline

- Motivation
- From 2D trajectories to 3D trajectories
- Application to Network Flow Modeling (NFD's)
 - Vehicular networks
 - Pedestrians and crowds
- Travel time reliability
 - Signature relations and trajectory data
 - Within-day and day-to-day variability
- Scenario-based approach to reliability modeling
 - Trajectory Processor for particle-based simulators
- Takeways, Limitations and Challenges

31

Complex interactions, Collective Effects

Relation between standard deviation of trip time per mile and mean trip time per unit distance



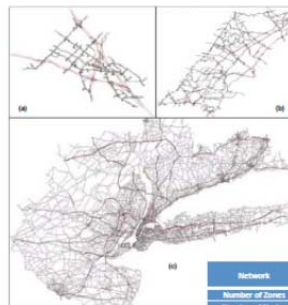
32

Theoretical Background – Travel Time Reliability

- Model has been validated and tested at different aggregation levels using different data sources (Mahmassani, H., Hou, T., Dong, J., TRB 2012)
- Data sources
 - Vehicle trajectories from simulation output
 - GPS probe data (location and time)
- Model works for different aggregation levels
 - Network level
 - O-D level
 - Path level
 - Link level

33

Simulated Trajectory Data

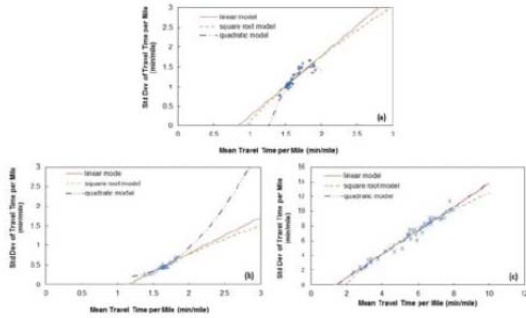


- Models are calibrated for different sizes of networks at different aggregation levels
- Three model forms are tested
 - Linear model
 - Square root model
 - Quadratic model
- Linear model gives best results
- Model parameters are estimated by Weighted Least Square (WLS) to accommodate heteroscedasticity

Network	Irvine	CHMET	New York City
Number of Zones	61	111	3697
Number of Nodes	326	2182	23406
Number of Links	438	3387	46490
Number of Vehicles	3035	15173	176632
Demand Duration [hr]	2	2	4

Simulated Trajectory Data

- Model comparison – network level analysis

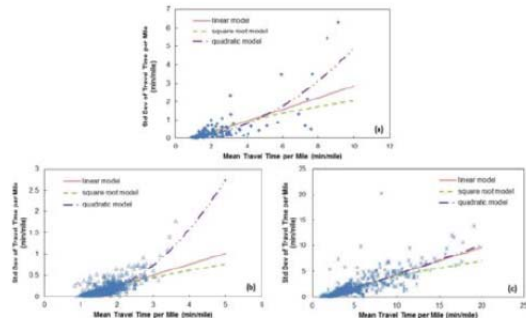


(a) Irvine; (b) CHART; (c) New York City

35

Simulated Trajectory Data

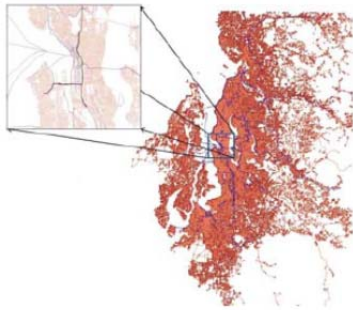
- Model comparison – path level analysis



(a) Irvine; (b) CHART; (c) New York City

36

GPS Probe Data

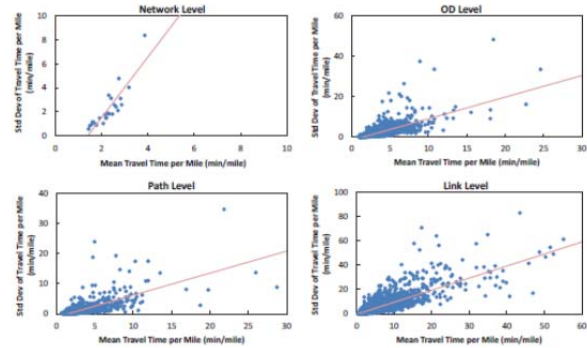


- Seattle network
 - ~600 zones
 - ~6000 nodes
 - 549,624 trips
 - ~400 participating vehicles

37

GPS Probe Data

- Seattle network

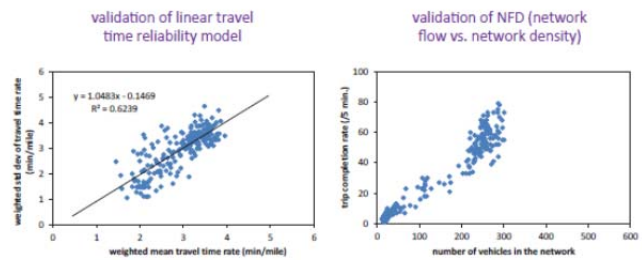


Validation by GPS Trajectory Data

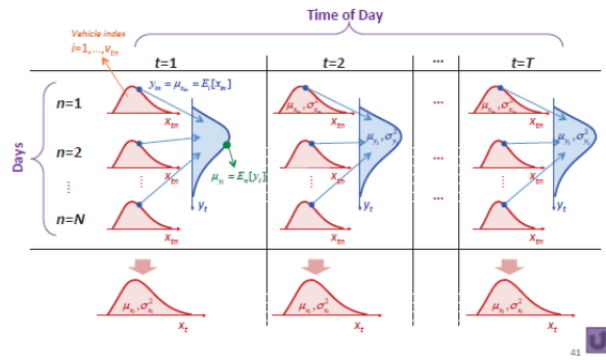


- NYC network
 - Vehicle trajectories collected by GPS devices
 - Two-week period from 2010/05/02 to 2010/05/17
 - ~10,000 trips are recorded on each day

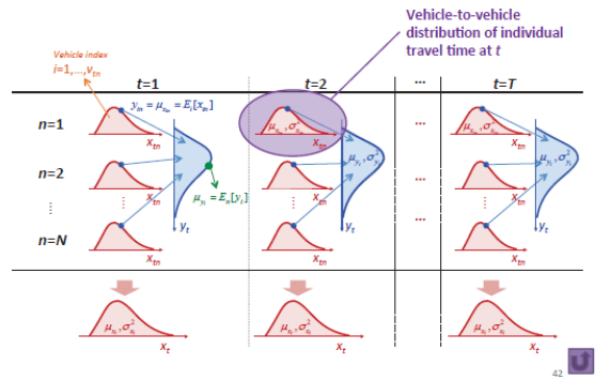
Validation by GPS Trajectory Data



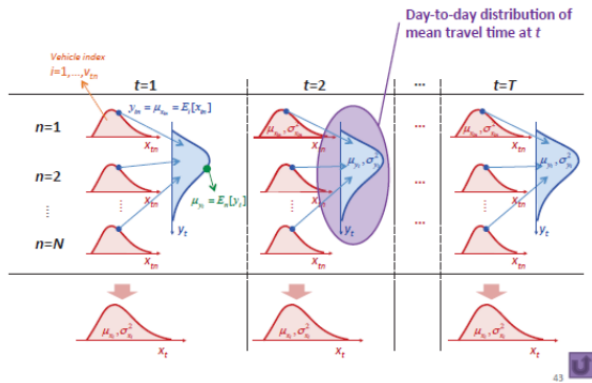
Characterizing Different Types of Travel Time Variability



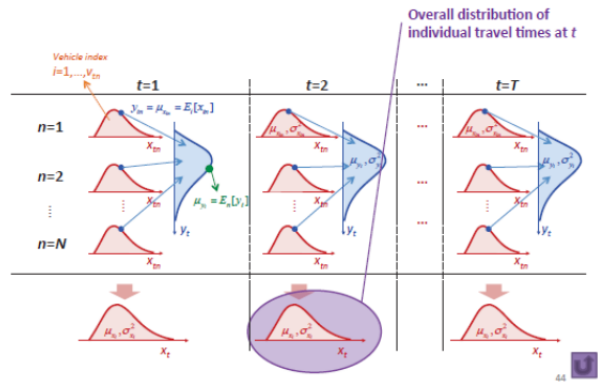
Characterizing Different Types of Travel Time Variability



Characterizing Different Types of Travel Time Variability



Characterizing Different Types of Travel Time Variability

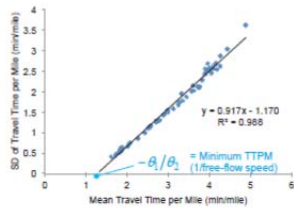


Characterizing *Vehicle-to-vehicle* Variability

- Linear Relationship between Standard Deviation (SD) and Mean
 - Jones et al. (1989) and Mahmassani et al. (2012, 2013)

$$\sigma_x = \theta_1 + \theta_2 \mu_x$$

- τ : travel time per unit distance; travel time per mile (TTPM)
- σ_x, μ_x : mean and SD of τ ; θ_1, θ_2 : coefficients



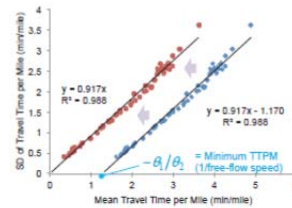
Characterizing *Vehicle-to-vehicle* Variability

- Travel Delay per Unit Distance
 - TTPM – minimum TTPM

$$x = \tau - (-\theta_1/\theta_2)$$

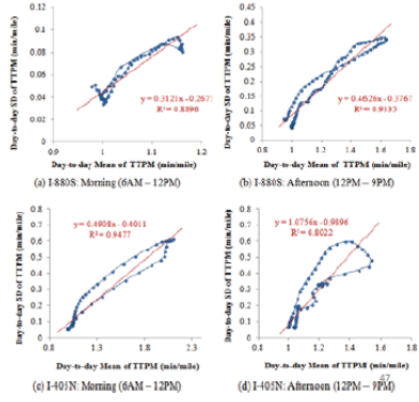
$$\Rightarrow \sigma_x = \theta_2 \mu_x$$

$$\begin{cases} E[x] = \mu_x = \mu_\tau + \theta_1/\theta_2 \\ SD[x] = \sigma_x = \sigma_\tau \end{cases}$$



Characterizing Day-to-day Variability

- Strong Correlation between SD and Mean
 - Herman and Lam(1974); and Richardson and Taylor (1978)
- Linear Relation between SD and Mean
 - May et al. (1989); Mazloumi et al. (2010); Yildirimoglu et al. (2013); Fosgerau (2010); and Fosgerau and Fukuda (2012)



Multiplicative Error Models

Vehicle-to-vehicle Distribution

Day-to-day Distribution

$$\sigma_{x_m} = \alpha \mu_{x_m}$$

$$\sigma_{y_t} = \beta \mu_{y_t}$$

$$x_t = y_t \varepsilon_x$$

$$y_t = \mu_{y_t} \varepsilon_y$$

$$\varepsilon_x \sim \text{Gamma}(\pi, 1/\pi)$$

$$\varepsilon_y \sim \text{Gamma}(\phi, 1/\phi)$$

$$SD[x_t] = SD[\varepsilon_x] y_t$$

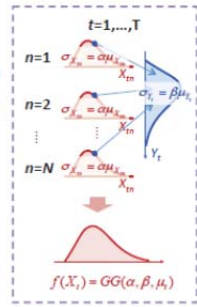
$$SD_n[y_t] = SD_n[\varepsilon_y] \mu_{y_t}$$

$$\alpha = 1/\sqrt{\pi}$$

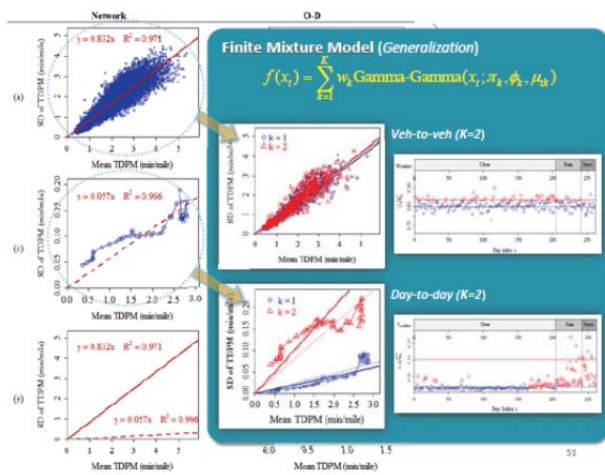
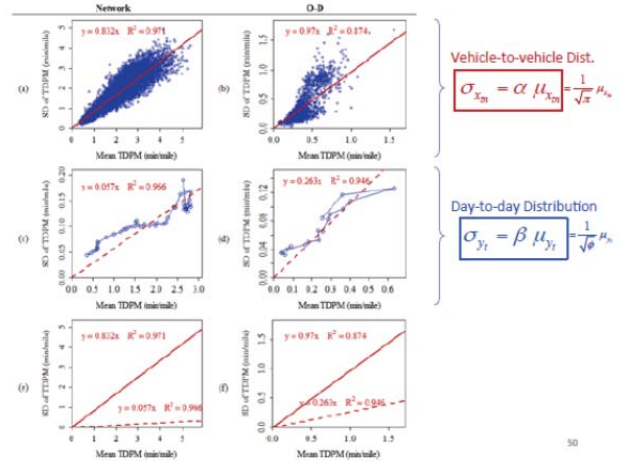
$$\beta = 1/\sqrt{\phi}$$

Gamma-Gamma Distribution

- Describe diverse scattering phenomena
 - target and clutter scattering in radar (Jakeman and Pusey, 1976; Lewinski, 1983);
 - irradiance fluctuations in optics (Al-Habash et al., 2001; Teich and Diamant, 1989);
 - reverberation in sonar systems (Gu and Abraham, 2001); and
 - fading and shadowing in wireless systems (Shankar, 2004)
- For Modeling Travel Time Variability
 - Shape π reflects veh-to-veh variability (i.e., $\alpha = 1/\sqrt{\pi}$: CV of individual travel delay across vehicles)
 - Shape ϕ reflects day-to-day variability (i.e., $\beta = 1/\sqrt{\phi}$: CV of daily mean level of travel delay across days)
 - Mean μ_x represents the mean level of travel delay at time t



$$E[X_t] = \mu_t \quad Var[X_t] = \mu_t^2 \frac{\pi + \phi + 1}{\pi\phi} = \mu_t^2 (\alpha^2 + \beta^2 + \alpha^2\beta^2)$$

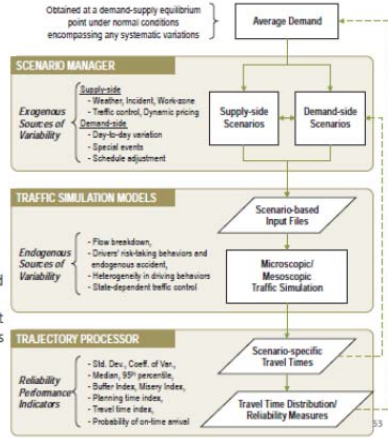


MPCormick
Northwestern Engineering
Northwestern University Transportation Center

Outline

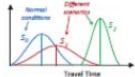
- Motivation
- From 2D trajectories to 3D trajectories
- Application to Network Flow Modeling (NFD's)
 - Vehicular networks
 - Pedestrians and crowds
- Travel time reliability
 - Signature relations and trajectory data
 - Within-day and day-to-day variability
- Scenario-based approach to reliability modeling
 - Trajectory Processor for particle-based simulators
- Takeways, Limitations and Challenges

Travel Time Reliability Analysis Framework

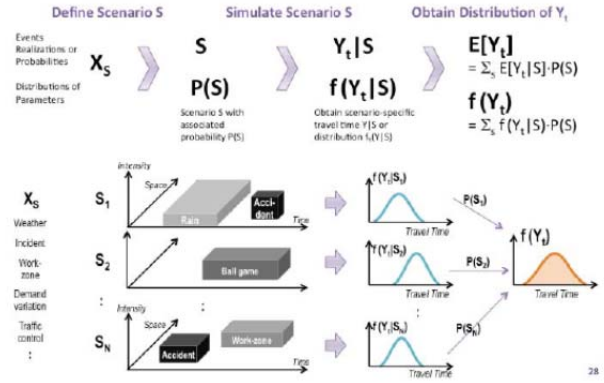


Trajectory Processor

Extract reliability-related measures from the vehicle trajectory output of the simulation models

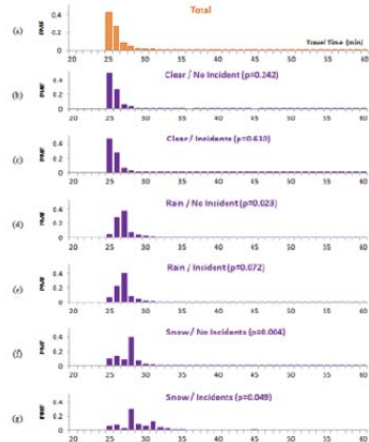


Scenario-based Reliability Analysis



Case Study

- Understand the impact of each scenario category
- Observe the overall travel time distribution



Vehicle Trajectory Processor

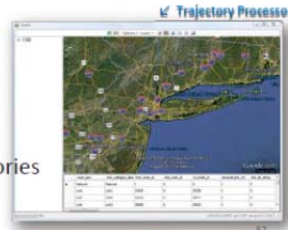
Main Role: Extract reliability-related measures from the vehicle trajectory output of the simulation models.



Trajectory Processor

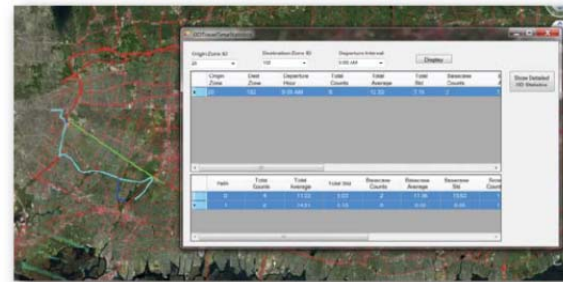
Application Main Functionalities

- Define **critical O-D pairs, paths and links** either via input text files or by selecting on the map
- From simulation outputs, for selected OD/path/links,
 - show **scenario-specific travel time distributions**
 - show **combined travel time distribution** (weighted by scenario probabilities)
 - extract various **reliability performance measures**
- Compare **simulated trajectories** with **observed trajectories** (e.g., TomTom GPS data)

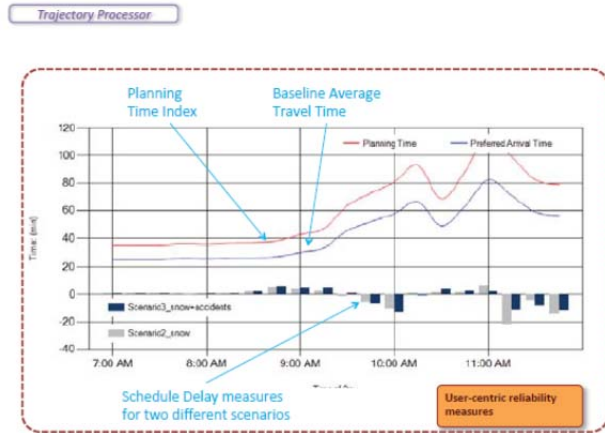
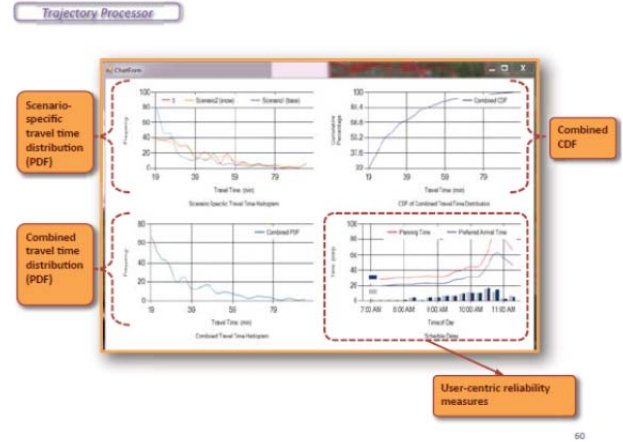
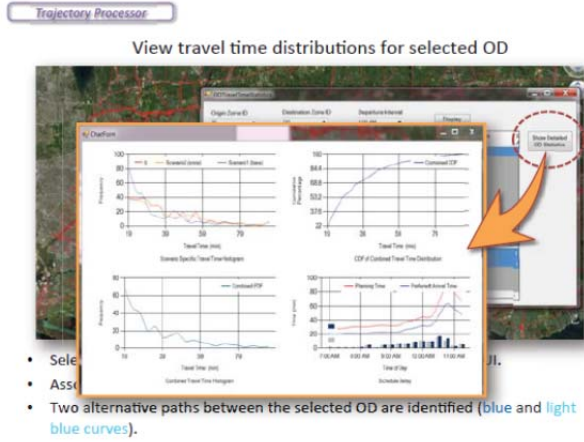


Trajectory Processor

Select O-D pair from the map



- Select two points on the map by clicking on the Google Earth GUI.
- Associated O-D pair is identified (**straight (green) line**).
- Two alternative paths between the selected OD are identified (**blue and light blue curves**).



2D Trajectories

Calibrate microscopic traffic relations (e.g. NGSIM)

Extract point measurements for mesoscopic and macroscopic model calibration and validation

3D Trajectories

MICROSCOPIC LEVEL

In addition to 2D capabilities, could model travel behavior choices

- route choice; responses to information, pricing, controls
- trip timing

Reliability characterization at individual vehicle level; both within day and across day variability

Ideal Revealed Preference manifestation

When based on smart phones or personal devices, not limited to car trips

When complete, allow capturing social influences and information

When coupled with transaction data enables wider range of behavioral responses

3D Trajectories

MESOSCOPIC and MACROSCOPIC

Retains ability to do same things as with point data

Time-dependent O-D demand estimation as input to dynamic network models

Takes validation to new levels, from patterns at points to spatio-temporal variation and user experience over entire travel; from facility-level to network-level.

Network-level relations

- Network fundamental diagram
- Reliability signature relations

Takeaways?

New era of trajectory-driven traffic and network performance analysis:

- More complete and compact description of system state
- Capture all aspects of individual actions (most complete record of actual behavior), with no loss of ability to characterize systems at any desired level of spatial and temporal aggregation/disaggregation
- Retain ability to extract stochastic properties of both individual behaviors and performance metrics
- Enable better model formulation/specification at all levels of resolution, and model calibration
- Most promising hope to recognize and capture collective effects and interaction mechanisms.

Limitations and Interesting Methodological Issues

- Partial trajectories
 - Censoring: incomplete trajectories of individual particles
 - Estimation of partially observed state variables: flow, density
 - Recognizing and correcting for selection bias
- Sampling trajectories
 - Sample designs and implications for estimation and model validation
- Non-uniqueness of underlying set of individual trajectories corresponding to observed aggregate (point-based) measurements
- Is it fair/reasonable to expect simulation tools to replicate individual vehicle trajectories, or only to replicate flow patterns in some aggregated fashion on links and
- Person trajectories, multimodal travel and activity-based modeling
- “Big data” aspects of very detailed trajectories
- ***Make trajectories your friend!***

65

THANK YOU!
QUESTIONS?

masmah@northwestern.edu

66

APPENDIX B
Paper Presentations

Automatic fitting procedure for the fundamental diagram

V. L. Knoop PhD
Delft University of Technology
Transport & Planning
Stevinweg 1
Delft, The Netherlands
+31 15 278 8413
v.l.knoop@tudelft.nl

W. Daamen PhD
Delft University of Technology
Transport & Planning
Stevinweg 1
Delft, The Netherlands
w.daamen@tudelft.nl

October 6, 2014

ABSTRACT

The fundamental diagram of a road, including free flow capacity and queue discharge rate, is very important for traffic engineering purposes. In the real world, most traffic measurements come from stationary loop detectors. This paper proposes a method to fit Wu's fundamental diagram to loop detector data. Wu's fundamental diagram is characterised by five parameters, being the free flow speed, wave speed, free flow capacity, queue discharge rate and jam density. The fits appear to be not very sensitive to the right value for the wave speed or the free flow speed. The proposed method therefore entails fixing these two parameters. The method consists of two steps. We first use a triangular fundamental diagram to separate the congested branch from the free flow branch. Then, the remaining three parameters of Wu's fundamental diagram are fitted on each of the branches using a least square fit. This method is shown to be robust for cases tested in real life, and hence very noisy, data.

INTRODUCTION

In traffic flow theory the fundamental diagram is an essential concept. The fundamental diagram relates two of the three variables average speed (v), flow (q) and density (k) to each other. If two of these variables are known, the third can be derived using the relation $q = kv$. Therefore, if only one variable is known, and the fundamental diagram is known, the traffic state can be determined.

Since Greenshields (1) showed the relation between speed and density, many theories have been developed to work with the fundamental diagram, including the often used LWR model (2, 3).

Although many papers are written on the shape of the fundamental diagram (see section 3.1), it is remarkable how much spread is present if one plots observed flows versus observed densities. This spread can have different causes, for instance:

1. The traffic is not stationary during the aggregation interval.
2. The traffic itself is heterogeneous.
3. There are detector failures, or other measurement errors.
4. Within the aggregation interval an integer number of vehicles is measured.
5. The speed is calculated as time mean speed instead of space mean speed.
6. Intrinsic measurement errors occur, for instance loops do not detect vehicles standing still (speed 0 km/h).

Although solutions have been found for several of these problems, most often traffic engineers still use loop detector data from freeways, with the abovementioned problems. If they want to apply theoretical concepts such as shock wave theory or the method of characteristics, they need to have a fundamental diagram. The parameters of the fundamental diagram should be calibrated for the observed road, as each road has its own characteristics, leading to a unique fundamental diagram.

Traffic control measures, such as ramp metering (4) or main lane metering (5) are exploiting the capacity drop, i.e. the difference between free flow capacity and the queue discharge rate. In order to find the right settings for these measures, it hence is required to find a reliable value for both values. For consistency of the results and calculation speed (if various locations are involved), an automatic fitting procedure is useful. However, due to errors in the data automatic fitting procedures often lead to extreme values for capacity or jam density, as will be shown later in the paper. This paper presents a method which can be applied in an automated way, preventing extreme values on noisy data – a property we will call robustness in the remainder of the paper. The key of the method proposed here (see also figure 4) is that first a triangular fundamental diagram is fitted to find the critical density, used to separate the free flow and the congested branch. Then, a more realistic fundamental diagram is fitted on the congested data and the free flow data separately. Variables which are known are fixed, thereby improving the robustness of the fit.

This paper first gives an overview of the literature on the fundamental diagram and the fitting procedures. Then, the experimental setup is introduced, showing the different choices in the research approach. This setup also includes the simulation setup and a description of the empirical data used to test the methodology. Then, the methodology is introduced and the simulation results are discussed, followed by a section presenting the fitting results on empirical data. Finally, we discuss the results and present the conclusions.

LITERATURE REVIEW

This section first describes the general shapes of the flow-density relationship, followed by a detailed discussion of the shape proposed by Wu. Then, the existing literature of fitting a functional form to this relationship is discussed.

Shapes of the fundamental diagram

Many shapes have been proposed for the fundamental diagram, ranging from very simple functions to functions with mathematically useful properties, functions including the microscopic traffic characteristics or relations based on empirical findings.

Greenshields (1) was the first to observe traffic flows and to hypothesize a linear relationship between speed and density. Applying the relation

$$q = kv \quad (1)$$

the linear relation between speed v and density k corresponds to a parabolic relation between speed v and flow q . The fundamental diagram defines a number of characteristics of the traffic flow. First of all, it determines the capacity, or the maximum flow that can be maintained for a short period of time. The corresponding density is the so-called critical density. This capacity distinguishes two states or regimes: for densities lower than the critical density, traffic is in an uncongested state, while for higher densities, traffic is in a congested state. The Greenshields fundamental diagram is called a univariate model, since both the uncongested regime and the congested regime are described using the same formula. Another well known univariate model is Drake's model (6), where the speed is an exponentially decreasing function of the density.

Two-variate models use distinct formulas to describe the congested and the uncongested regime. The simplest two-variate model is a triangular fundamental diagram, that describes both regimes with a straight relationship. The slope of the uncongested regime is equal to the free speed, while the slope of the congested regime equals the shock wave speed (describing the speed with which the congestion front is moving upstream). The latter shows a direct relation between the fundamental diagram and shockwave theory (2). Daganzo (7) introduced the truncated triangular fundamental diagram. Here, the capacity is not only reached at the critical density, but for a certain range of densities the flow is constant, and maximal. Edie (8) was among the first researchers to show that a discontinuous relation might be more appropriate to describe traffic dynamics. He distinguished the regime of the free traffic, with the so-called free flow capacity as maximum flow, and the congested traffic, having the so-called queue discharge rate as maximum flow. As the curve is discontinuous, free flow capacity is higher than the queue discharge rate. This is called the capacity drop (9). Koshi et al. (10) introduced an inverse lambda shaped fundamental diagram, which included this capacity drop. In this diagram, traffic has a free speed up to a capacity point. However, the congested branch does not start at capacity, but connects below the capacity point to the free flow branch. An addition to the inverse lambda shaped fundamental diagram is made by Wu (11). He assumes that the speed in the free flow branch is not constant, but decreasing with increasing density. The shape of the free flow branch is determined by the overtaking opportunities, which depend on the number of lanes. Figure 1 visualises the density-flow and density-speed diagram mentioned above.

Kerner (12) proposed a different traffic flow theory, the so-called three phase traffic flow theory. Instead of distinguishing an uncongested regime and a congested regime, Kerner describes three phases of traffic. In addition to the uncongested regime, the congested regime has been split

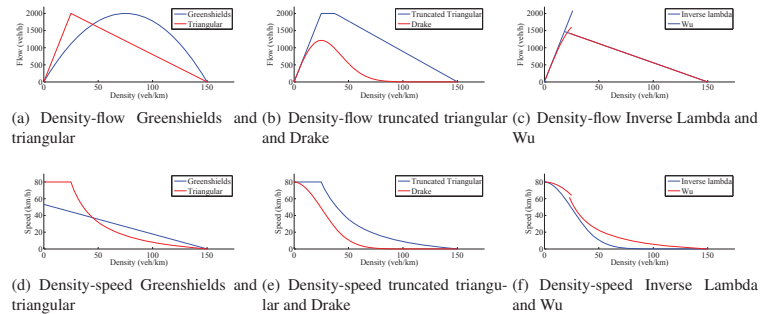


FIGURE 1 Different shapes of the fundamental diagram

into two distinct phases: synchronised flow and wide moving jam. In a wide moving jam traffic stands still and the wave moves upstream (similar to stop-and-go waves); in synchronised flow, the traffic is moving. Contrary to other researchers, Kerner claims that in this state several equilibrium distances exist for the same speed, leading to an area rather than a line in the flow-density plane.

So far, the described fundamental diagrams are all deterministic. Recent research (13) has tried to explicitly include stochastic aspects in the fundamental diagram.

Each of the fundamental diagrams mentioned above has its advantages and disadvantages. The existence of a capacity drop requires the fundamental diagram to have at least two regimes: a free flow regime and a congested regime. In free flow, the most realistic model is that speed decreases with increasing densities, implying a non-linear free flow branch. Taking these two requirements as a starting point, we have chosen to fit Wu's fundamental diagram to our data.

Wu's fundamental diagram

The fundamental diagram by Wu (11) is based on the overtaking possibilities. Those differ for roads with a variable number of lanes. Wu (11) defines a fundamental diagram by prescribing the speed as function of density. He argues that the speed changes gradually from the free flow speed v_{free} to the critical speed v_{crit} . The way it decreases depends on the number of lanes, N . In an equation, this gives:

$$v_{free \text{ flow branch}}(k) = v_{free}(v_{free}v_{crit}) \left(\frac{k}{k_{crit}} \right)^{N-1} \quad (2)$$

Multiplying both sides with the density gives the relation between the flow (speed times flow is density: equation 1) and the density. For the congested branch, Wu proposes a straight line in the flow-density relationship, starting at a lower flow than the free flow, hence incorporating a capacity drop. It can be easily verified that these equations can be rewritten, and different parameters can be chosen to represent the same shape. For this paper, we rewrote the equations such that the parameters of the fundamental diagram were as meaningful as possible. Note that rewriting the equations or changing variables does not change the relationship, nor the parameter values. Table 1 shows the variables and their default values being used for this paper.

TABLE 1 The parameters of Wu's fundamental diagram and their default values

name	symbol	default value
Free flow speed	v_{free}	115 km/h
Wave speed	w	-18 km/h
Free flow capacity	C_f	2400 veh/h
Queue discharge rate	C_q	1800 veh/h
Jam density	k_{jam}	150 veh/km/lane

Algorithms to fit a fundamental diagram

Although the shape of the fundamental diagram is under discussion, not much literature can be found on how to estimate a fundamental diagram. More researchers discuss which data to use for such an estimation.

Typically, fundamental diagrams are estimated from aggregated data at a specific location (14, 15). Depending on the type of detectors used to derive these data, occupancy or densities are observed. In addition, speeds can be averaged over time or harmonically averaged. One of the drawbacks is the existence of transient phases, which are difficult to distinguish from equilibrium phases. Only the latter should be used to derive the fundamental diagram, as this describes traffic in an equilibrium state. When using the fundamental diagram to predict how traffic states propagate (both in time and in space), the estimation should be based on spatial measurements (in accordance to Edie's definition of equilibrium). As detectors only collect data at a single cross-section, probe vehicles can be used (16). Chiabaut et al. (17) use the passing rate to estimate a fundamental diagram, as this is independent of the traffic flow state. Using NGsim data, they prove that a linear fundamental diagram can be used to describe the traffic and that both the congested wave speed and the jam density can be accurately predicted.

Although literature clearly has shown drawbacks of detector data, this is the type of data typically available. Therefore, we have developed a method to estimate the fundamental diagram based on detector data.

EXPERIMENTAL SETUP

This section introduces the different parameters to be fit and which combinations of these parameters we compare. Also, the setup of the simulation to be used to assess the methodology is elaborated upon, and some details are given on the empirical data we use for evaluation.

Influencing elements in fitting a fundamental diagram

In the process of the fitting of the fundamental diagram several aspects play a role. We can fix parameters, that is, not all the parameter values need to be estimated. Also, different speed averages might lead to different results. Moreover, the stochasticity of real traffic might be different from a traffic simulation. This section comments on these variables, while in the next section it is indicated which combinations are tested.

Fixing parameters

Fixing one or more parameters will increase the robustness of a fit, simply because one can set the parameter to a value which must be approximately right. Thereby, it is avoided that a variable is set to an extreme value because another variable is wrong.

Wu's fundamental diagram has five parameters, of which 2 are the capacities which are to be estimated. The remaining 3 are the free flow speed, the wave speed and the jam density. The

TABLE 2 The analyses performed

Par fixed	Data source	Type of data	Results in
-	Simulation	Detector, arithmetic mean	figure 7(a)
v_{free} and w	Simulation	Detector, arithmetic mean	figure 7(c)
v_{free}	Simulation	Detector, arithmetic mean	figure 7(d)
w	Simulation	Detector, arithmetic mean	figure 7(e)
-	Simulation	Edie	table 3
v_{free} and w	Simulation	Edie	table 3
-	Simulation	Detector, harmonic mean	figure 7(f)
v_{free} and w	Simulation	Detector, harmonic mean	figure 7(g)
w	Simulation	Detector, harmonic mean	figure 7(h)
-	Real	Detector, arithmetic mean	figure 7(i)
v_{free} and w	Real	Detector, arithmetic mean	figure 7(j)
w	Real	Detector, arithmetic mean	figure 7(k)

jam density is most constant, and can be determined by other methods as well (e.g., (17, 18)). Also the free flow speed can relatively easily be estimated based on the speed limit. The jam density is more difficult to estimate on beforehand, and this will not be considered in the tests here. Fixed wave speed and/or free flow speed will be considered.

Type of data in relation to speed averaging

Some of the problems when determining the fundamental diagram are caused by the fact that the reported mean speeds are time mean speeds. For the density calculation by $k = q/v$ (eq. 1) one needs the space mean speed, or as approximation the harmonically averaged speed. Figure 6(a) shows the differences of these different density calculations for the simulations. Still, high density areas hardly occur, but the congested conditions are better represented in the diagram. This has to do with the fact that the congested states are more affected by the speed averaging procedure (19).

Another cause of problems when determining the fundamental diagram is the fact that the measurements are taken at one location. From the simulation, also the full trajectories are known. That means that we can apply Edie's definitions (8). In particular, we divide the road into segments of 100 meters and consider time intervals of 10s, leading to an area in space and time $A = 100 * 10 = 1000$ ms. For these rectangles in space and time, the total distance travelled D by all vehicles and the total time spent in the area T can be calculated. The flow and density then are calculated by $q = D/A$ and $k = T/A$. In this paper we will look at the consequences of using detector data or Edie's definitions based on the full trajectories.

Real world data or simulation data

In the paper, first the method is applied to simulation data. The advantage of using data from a microscopic simulation program is that the ground truth (i.e., the underlying fundamental diagram) is known, and there are no measurement errors. Moreover, all type of data (time mean speeds, harmonically mean speeds, Edie's generalised density using trajectories) can be obtained from the simulation. The simulation setup is chosen simple in order to test the properties of the proposed methodology. Later, the same method is applied to real world data.

Analyses

The section above indicated several combinations of influencing elements in the fitting procedure. Not all combinations are tested in this paper. Table 2 shows the analyses we chose for this paper.

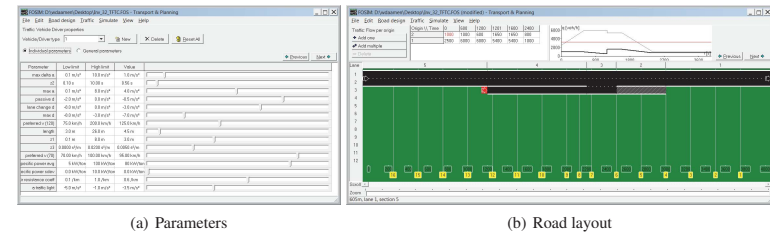


FIGURE 2 Fosim simulation setup

The selection we made is based on a combination of simulation data and real data, but the emphasis is on simulation data. The main reason is that the ground truth (theoretical fundamental diagram) is known for these simulation data. Moreover, we can get arithmetic mean speeds and harmonic mean speeds at (virtual) detectors as well as Edie's data from the trajectories. So, in short, all information is available to perform extensive tests on the proposed method.

As said, fixing of parameters might increase the robustness of the fitting results, so we will consider fixing the free flow speed and/or the wave speed. As will turn out, especially fitting the congested part is difficult, particularly if only the arithmetic mean speed is known. Therefore, for real data only fixed wave speed is considered.

Simulation

In this section we discuss the various elements of the simulation, such as the simulation model and the setup of the simulation.

Simulation program

The methodology is tested using simulation data. The advantage of traffic simulation is that the ground truth is known (see figure 3(a)), although lane changing effects are not taken into account. To generate simulation data we use Fosim (20). Fosim is the only simulation model that has been validated for Dutch freeways, and it is used by the Dutch road authority for ex-ante capacity assessments for new infrastructure (21).

Fosim implements a Wiedeman principle (22) in the car-following model. The equilibrium spacing drivers prefer is a quadratic function of their speed:

$$s = z_3 v^2 + z_2 v + z_1 \tag{3}$$

Fosim implements by default two user classes: passenger cars and Heavy Good Vehicles (HGV). For passenger cars, three different types of drivers are distinguished: aggressive drivers, more timid drivers and drivers driving vans. For each of these driver-vehicle combinations, input parameters are determined (see the screenshot in figure 2(a)).

Simulation setup

For the sake of simplicity and traceability, we only consider a homogeneous flow of timid drivers in passenger cars in the simulation. We have used the default (thus validated) parameters for this

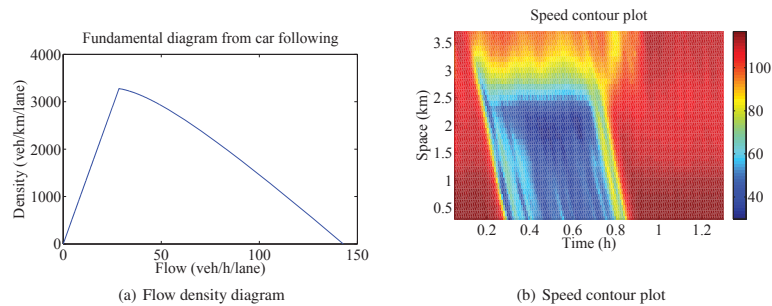


FIGURE 3 The traffic characteristics in Fosim simulation model

vehicle class. For our driver population, this means that the parameter values for z_1 , z_2 , and z_3 are respectively 3m, 0.72s and $0.0050s^2/m$. This, together with the vehicle length of 4m gives the speed spacing diagram, which can be translated into a flow-density diagram (see figure 3(a)). Note that this theoretical calculation does not take stochasticity into account. The resulting capacity from this relationship seems to be rather high, at a value of approximately 3275 veh/h/lane, equalling a headway of 1.1s. This is the theoretical capacity in case all vehicles are following (in both lanes). These flow values are being observed in the Netherlands for passenger cars only in the left lane. For instance, on the A4 section between The Hague and Leiden (see the section on fitting results on empirical data) we find a maximum flow in the left lane of 56 vehicles in one minute, equalling 3360 veh/h.

The case study we consider for estimating the fundamental diagram is as simple as possible. We consider a two lane freeway with an onramp, see figure 2(b). The top of this figure shows the demands, both for the main road (origin 2), and for the onramp (origin 1). With these demands the capacity of the road is temporarily exceeded. First, a stop-and-go wave is created by increasing the flow on the main road, where the head of the congestion is just downstream of the end of the onramp. Secondly, bottleneck congestion is created by a higher inflow from the ramp and a lower inflow from the main road.

We have collected speeds, flows and density at detectors, located 250m apart, see the yellow boxes in figure 2(b). As these detectors observe passing moments of vehicles per lane, only data from the two main lanes are used for the estimation of the fundamental diagram. As merging occurs at the merging lane, causing disturbance of the process, we have excluded detectors 6 and 7.

In Fosim, simulations often lead to congestion where vehicles come to a complete stop. Also, passenger cars accelerate at a high rate, thus not passing detectors at a low speed. As mentioned, when traffic is completely stationary (standing still), no vehicles pass the detector, and no traffic is observed. So the only traffic that is being observed is the relatively fast moving – because quickly accelerating – traffic. To overcome these problems, we create other types of bottlenecks as well. Instead of creating an infrastructure bottleneck, we halve the free speeds. This way, we

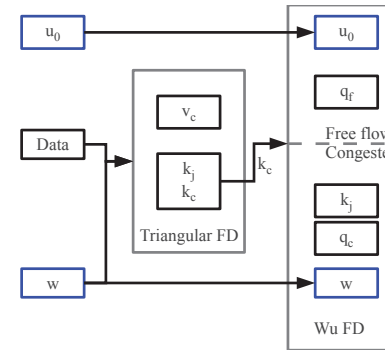


FIGURE 4 The overview of the fitting procedure

prevent vehicles from standing still, and thus we prevent the introduction of model bias into our data set.

Experimental data

The method proposed in this paper is also applied on data from the A4 freeway in the Netherlands between Amsterdam and Rotterdam. The stretch we consider is 10km long (from Leiden to The Hague). On this freeway only loop detector data are being collected, consisting of flows and time mean speeds.

FITTING METHODOLOGY

The main innovation of the method proposed in this paper is that we first use a triangular fundamental diagram to separate the congested branch from the free flow branch. Then, Wu’s fundamental diagram is fitted on each of the branches. A graphical overview of the method is presented in figure 4. In this fitting process, some variables can be fixed. In this figure, it is indicated that the wave speed and the free flow speed are taken fixed, which will turn out to give the most robust fitting results.

Use of data points

We assign a weight p to each of the data points based on their reliability. This weight is not constant, as data points that can be trusted should contribute more to the quality of the fit than unreliable points and outliers. The following weights are assigned:

- Speeds under 10 m/s cannot be trusted. It would be very rare to have situations where the speed is constantly 10 km/h. Therefore, this will be a mixture of states, in which only the moving vehicles are measured. Due to the nature of local measurements, vehicles standing still will not be measured at all. Moreover, the approximation of the space mean speed by time mean speed will be worse when the spread of (measured) speeds is higher. Therefore, we give measurements with these low speeds a lower weight. In order to get a

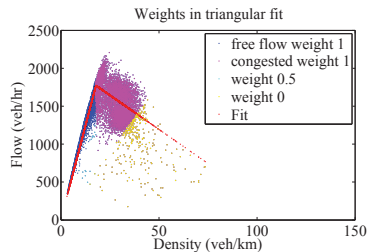


FIGURE 5 The triangular fit

robust fit, we do not want to pre-specify the speed which is too low, but we would like to relate that to the speeds resulting from the fundamental diagram. If the weight would be zero, the fitting procedure benefits from choosing a very high free speed, thereby ignoring all aggregation intervals with a lower free speed. Since this is undesirable, a weight of 0.5 is given to these points.

- The measurements are taken at a freeway. Therefore, in low density conditions we expect high speeds. Consequently, low speeds at low densities are a measurement error. This could be due to the fact that speeds are time mean speeds, which distorts the speed and hence the estimation of density. This means these data points are unreliable. If the speeds are below 20 m/s in free flow, the points are neglected.

Performance measures and optimization

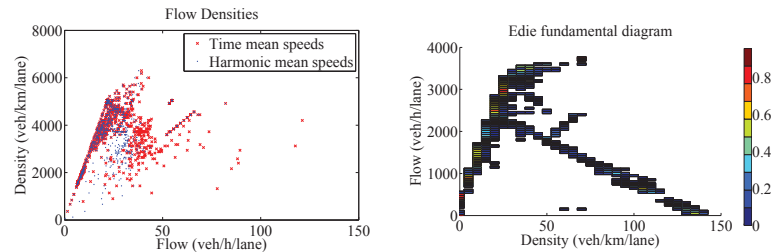
The fit will give a predicted flow (\hat{q}) based on the measured density k_i . The error for a data point (one observation of flow and density) can be calculated as the difference between the observed flow q and the predicted flow \hat{q} . As error of the fit we consider the weighted root mean square error of all data points, indicated by ϵ :

$$\epsilon = \sqrt{\frac{\sum_i p_i (\hat{q}_i - q_i)^2}{\sum_i p_i}} \tag{4}$$

The goal of the optimization is to minimize this error.

Fitting a triangular fundamental diagram

First, we will estimate the triangular fundamental diagram. This will give a robust estimation of the critical density. That critical density will be used later on to separate the free flow branch from the congested branch. The separation of the two branches is made based on density. A separation on speed is less reliable. The fitting procedure could end up fitting a high free flow speed, which fit data points with the high speeds very well. The other points, which one would assign to the free flow branch, would be moved to the congested branch, which has a higher spread anyway, and therefore would fit in the line of expectation.



(a) The points in the fundamental diagram for time mean speeds and harmonically averaged speed (b) Fundamental diagram according to Edie's definitions

FIGURE 6 Observations of combinations of speed and density using different (speed) definitions

This fundamental diagram has three degrees of freedom: the free flow speed, the critical density and the jam density. Note that the capacity is determined by the free flow speed times the critical density. In our approach the shock wave speed is fixed, so two degrees of freedom remain: the free flow speed and either the critical density or the jam density, which is for fitting purposes equivalent.

The result of this fitting procedure is a triangular fundamental diagram. This triangular fundamental diagram will give the critical density which is input to the fitting procedure of the Wu fundamental diagram (below).

Fitting Wu's fundamental diagram

For the same reasons as above, we assume a fixed shock wave speed w . Moreover, for freeway driving, we can assume a fixed free speed. This is because it does hardly influence capacity, and it is relatively constant for freeways. In this paper, we set the free flow speed u_0 to 30 m/s.

The critical density is used to split the data into free flow points and congested data points. There might be an error in the estimation of the critical density. Therefore, data points with a density between 90% and 120% of the estimated critical density from the triangular fit are not taken into account. If one considers the fundamental diagram, this is generally not an issue: the fitting of the diagram can be done on the "slopes" of the fundamental diagram. The points near the top of the diagram can belong to both branches, thus introducing an error in the fit, but these points are not necessary to find the slopes of the congested branch and the free flow branch respectively.

The remaining parameters of the fundamental diagram which need to be estimated are: (1) the free flow capacity, (2) the queue discharge rate and (3) the jam density (see figure 4).

SIMULATION RESULTS

This section presents the results of the fitting procedure. We start showing the fundamental diagram using Edie's definitions. Then, the estimation results when different parameter values are fixed in order to see the sensitivity of each parameter are shown. Finally, the improvement of the fit when using harmonic mean speeds is identified.

TABLE 3 The value of the parameter estimates

Fixed parameters	Detector (arithm. mean speed)				Eddie				Harmonic mean				
	-	$v_{free}&w$	v_{free}	w	readout	-	$v_{free}&w$	-	$v_{free}&w$	w	-	$v_{free}&w$	w
v_{free} (km/h)	126	[115]	[115]	121	100	116	[115]	126	[115]				
w (km/h)	-4.6	[-18]	-8.2	[-18]	-22	-3.6	[-18]	-5.7	[-18]	[-18]			
C_f (veh/h/lane)	2789	4015	3880	3311	3400	4919	5942	2949	4126	3264			
C_q (veh/h/lane)	2039	2149	2069	2098	2250	2460	2593	2169	2286	2258			
k_{jam} (veh/km/lane)	477	144	276	144	135	709	171	410	153	153			

Eddie's definitions

The fundamental diagram using the points obtained by Eddie's definitions can be visualized by the proportion of observations for a combination of density and flow. This is done in figure 6(b). The line of the fundamental diagram agrees to a large extent with the theoretical lines; however, the diagram constructed from the trajectories shows a large capacity drop, while this is not present in the theoretical curve. This is because the theoretical curve is only based on car-following behavior, and does not take into account that drivers might leave a larger gap than desired when leaving the queue.

Fix various parameters

This section describes what happens to the fit of the fundamental diagram if some of the parameters are fixed, while applying the proposed methodology. Table 3 presents the results for the different parameters values, as well as the values for these parameters as derived from the graph using Eddie's definitions. For Eddie's definitions, we read the parameter values for the ground truth from the figure; this is possible because the figure give very clear lines.

The table shows large differences between the different fits. The graphical representation is also shown in figure 7. Figure 7(a) shows the curve if no parameters are fixed. The jam density is estimated at 477 veh/km/lane, which is much higher than the approximately 145 veh/km/lane which follows from theory and the analysis of trajectories using Eddie's definitions. Figure 7(b) shows the fundamental diagram with the free flow speed and the wave speed being fixed. The shape and parameters matches relatively well with the shape of fundamental diagram from Eddie's definitions, although the free flow capacity is higher. In fact, it seems that in the fundamental diagram according to Eddie's definitions the speed decreases strongly just before the free flow capacity point is reached. Therefore, a comparable fit of the shape of Wu leads to a higher capacity point. Given these intrinsic shortcomings, the fit is reasonably well.

It is of course a disadvantage to fix both the free flow speed and the wave speed. Therefore, it is also checked whether the same results can be obtained by fixing one of these. Figure 7(d) shows that fixing only the free flow speed does not alleviate this problem, but figure 7(e) shows that fixing only the wave speed does.

Effect of the speed averaging

As mentioned before, speed averaging has an effect on the points in the fundamental diagram, particularly on the congested branch. Figure 7(f) shows the fit of the flow-density diagram where the density is computed using the harmonic mean speed and all parameters are estimated. In figure 7(g) the free flow speed and the wave speed have been fixed in the fit. Overall, the density can be better estimated when the density is based on the harmonic mean speed. However, as the amount of points along the congested branch is very limited, not much information is known on

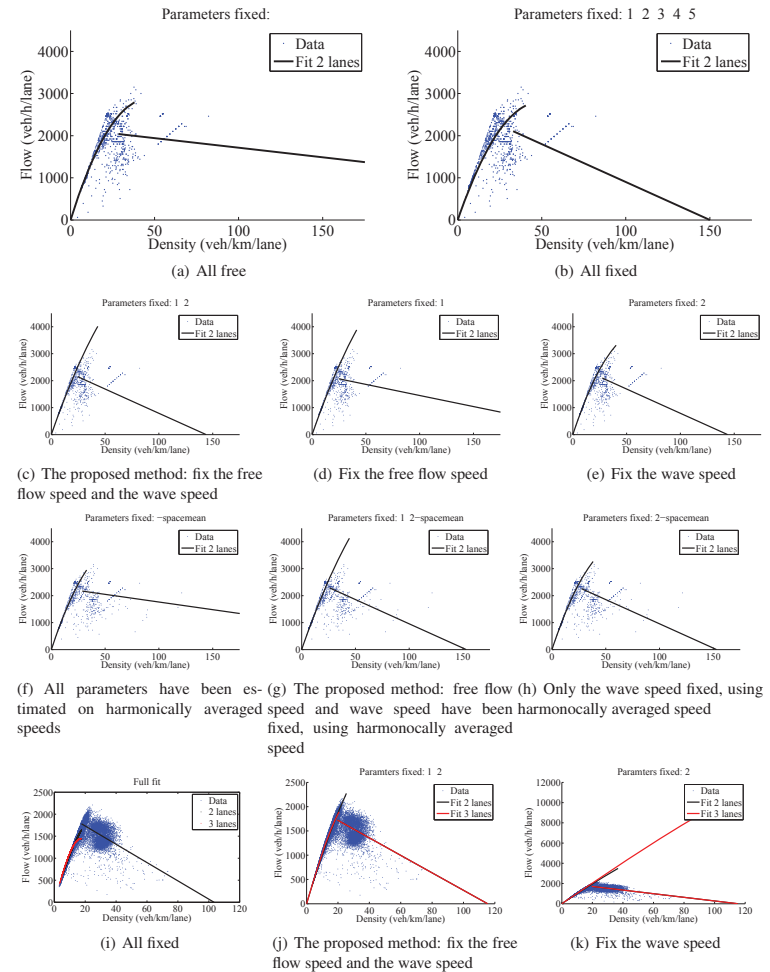


FIGURE 7 Fits of the fundamental diagrams on simulation data (a-h) and real data (i-k)

the shape of the congested branch. This is due to the very low speeds at which these points are located. The fit with all 5 parameters estimated still has a very large estimated jam density of 410 veh/km/lane, hence this method is unfeasible for practice. This can be improved by fixing the wave speed, and possibly also the free flow speed. Especially the method only fixing the wave speed yields good results (see table 3), with a closely matching free flow capacity and queue discharge rate.

APPLYING THE METHOD TO EMPIRICAL DATA

Figure 7(j) shows the fitting results on the real data. First, figure 7(i) shows what happens if all five parameters of Wu's fundamental diagram are fitted simultaneously. This leads to a fit which visually does not seem to follow the data points. Moreover, the free flow capacity is estimated to be (much) lower than the queue discharge rate. If we fix the wave speed and the free flow speed (see figure 7(j)), the fit follows the expected shape of the fundamental diagram better. However, the capacity values are slightly lower than the capacities indicated in the Dutch version of the Highway Capacity Manual (23). The fitted capacities are 1890 veh/h/lane for the three lane section and 2270 veh/h/lane for the two lane section, where the Dutch Highway Capacity Manual indicates values of 2100 veh/h/lane for both situations. Also the data points seem to point at a slightly higher capacity. However, for design purposes, a slight underestimation of the capacity is better than an overestimation. The queue discharge rate is estimated at 1725 veh/h/lane.

Fixing only the wave speed, which gives a good fit with simulation results, is not working for the case with real data, since the free flow capacity will be given an extreme value (see figure 7(k)).

We can thus conclude that the method fixing the wave speed and the free flow speed can also be used to fit a fundamental diagram on empirical data.

CONCLUSIONS

This paper presents a method to fit a fundamental diagram to traffic data, thereby estimating the free flow capacity and queue discharge rate. Difficulty is that points near the capacity can be assigned to either of the two branches, which makes the fitting process difficult, leading to extreme values. We propose to first separate the free flow branch from the congested branch by a triangular fit. Then, the free flow speed and the wave speed are fixed, thereby limiting extreme values. The method proved to yield reasonable, though not perfect, results on simulation data. Also for more noisy real world data, the method gives realistic capacity estimates.

ACKNOWLEDGEMENT

This research was sponsored by the NWO grant "There is plenty of room in the other lane", as well as the NWO Aspasia grant of Daamen.

References

- [1] Greenshields, B. D., A Study of Traffic Capacity. *Proceedings Highway Research Board*, Vol. 14, 1934, pp. 448–477.
- [2] Lighthill, M. J. and G. B. Whitham, On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads., *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 229, No. 1178, 1955, pp. 317 – 345.

- [3] Richards, P. I., Shock waves on the highway. *Operations Research* 4, Vol. 4, 1956, pp. 42 – 51.
- [4] Papamichail, I., M. Papageorgiou, V. Vong, and J. Gaffney, Heuristic Ramp-Metering Coordination Strategy Implemented at Monash Freeway, Australia. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2178, 2010, pp. 10–20.
- [5] Carlson, R. C., I. Papamichail, M. Papageorgiou, and A. Messmer, Optimal Motorway Traffic Flow Control Involving Variable Speed Limits and Ramp Metering. *Transportation Science*, Vol. 44, No. 2, 2010, pp. 238–253.
- [6] Drake, J. S., J. L. Schöfer, and A. May, A Statistical Analysis of Speed Density Hypotheses. In *Proceedings of the Third International Symposium on the Theory of Traffic Flow* (L. C. Edie, R. Herman, and R. Rothery, eds.), Elsevier North-Holland, New York, 1967.
- [7] Daganzo, C. F., *Fundamentals of Transportation and Traffic Operations*. Pergamon, 1997.
- [8] Edie, L., Discussion of traffic stream measurements and definitions. In *Proceedings of the Second International Symposium on the Theory of Traffic Flow*, OECD, Paris, France, 1965.
- [9] Banks, J., Review of Empirical Research on Congested Freeway Flow. *Transportation Research Record, Journal of the Transportation Research Board*, No. 1802, 2002, pp. 225–232.
- [10] Koshi, M., M. Iwasaki, and I. Ohkura, Some findings and an overview on vehicular flow characteristics. In *Proceedings of the 8th International Symposium on Transportation and Traffic Theory*, Univ. of Toronto Press, Toronto, 1981, pp. 403–426.
- [11] Wu, N., A new approach for modeling of Fundamental Diagrams. *Transportation Research Part A: Policy and Practice*, Vol. 36, No. 10, 2002, pp. 867–884, doi: DOI: 10.1016/S0965-8564(01)00043-X.
- [12] Kerner, B. S., *The Physics Of Traffic: Empirical Freeway Pattern Features, Engineering Applications, And Theory*. Springer, Berlin, 2004.
- [13] Wang, H., D. Ni, Q. Chen, and J. Li, Stochastic modeling of the equilibrium speed-density relationship. *Journal of Advanced Transportation*, Vol. 47, 2013, pp. 126–150.
- [14] Cassidy, M., Bivariate relations in nearly stationary highway traffic. *Transportation Research Part B: Methodological*, Vol. 32, No. 1, 1998, pp. 49–59.
- [15] Kockelman, K., Changes in flow-density relationship due to environmental, vehicle, and driver characteristics. *Transportation Research Record*, Vol. 1644, 1998, pp. 47–56.
- [16] Neumann, T., P. Bohnke, and L. Touko Tcheumadjeu, Dynamic representation of the fundamental diagram via Bayesian networks for estimating traffic flows from probe vehicle data. In *2013 16th International IEEE Conference on Intelligent Transportation Systems - (ITSC)*, 2013.
- [17] Chiabaut, N., C. Buisson, and L. Leclercq, Fundamental Diagram Estimation through Passing Rate Measurements in Congestion. In *Proceedings of 87th Annual Meeting of the Transportation Research Board*, Washington D.C., 2008.
- [18] Leclercq, L., Calibration of flow-density relationships on urban streets. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1934, No. 1, 2005, pp. 226–234.
- [19] Knoop, V. L., S. P. Hoogendoorn, and H. J. Van Zuylen, Empirical Differences between Time Mean Speed and Space Mean Speed. In *Proceedings of Traffic and Granular Flow 07*, Springer, Paris, France, 2007.
- [20] Dijkster, T. and P. Knoppers, *Fosim 5.1, Gebruikershandleiding*. Delft University of Technology, 2006.

- [21] de Leeuw, M. and T. Dijkstra, *Praktijkstudies Fosim 4*. Delft University of Technology, 2000.
- [22] Wiedemann, R., *Stimulation des Strassenverkehrsflusses*. Heft 8 der Schriftenreihe des ifv, Universität Karlsruhe, 1974.
- [23] Dutch Road Authority, *Capaciteitswaarden Infrastructuur Autosnelwegen*. Rijkswaterstaat, 2011.

ACCURACY OF PEDESTRIAN AND TRAFFIC FLOW MODELS: MEANINGFUL QUANTIFICATIONS

Femke van Wageningen-Kessels (corresponding author)
Delft University of Technology
Stevinweg 1, 2628CN Delft, The Netherlands
+31 6 2848 5258 / +968 9325 1057
f.l.m.vanwageningen-kessels@tudelft.nl

Winnie Daamen
Delft University of Technology
Stevinweg 1, 2628CN Delft, The Netherlands
+31 15 27 85927
w.daamen@tudelft.nl

Serge Hoogendoorn
Delft University of Technology
Stevinweg 1, 2628CN Delft, The Netherlands
+31 15 27 85475
s.p.hoogendoorn@tudelft.nl

July 25, 2014

ABSTRACT

Accuracy measures are used to quantify the accuracy of road traffic and pedestrian flow observations, models, simulations and predictions. We propose two new ways to measure the accuracy, by introducing and quantifying the phase error and the diffusion error. These errors take into account specific features that are of interest in traffic and pedestrian flow models. The phase error penalises a wrong estimate of the location of a high density region, the diffusion error penalises too smooth transitions between high density and low density regions. Combining the results of both accuracy measures, gives better insight than traditional accuracy measures into how the methodologies or models may be improved. Test cases show the application of the accuracy measures for the selection of a simulation method and for parameter estimation. Moreover, they show how the measures can be applied both for (one dimensional) road traffic flow and for (two dimensional) pedestrian flow.

INTRODUCTION

Practical applications of traffic and pedestrian flow theory include many steps in which accuracy is important: the flow is observed, a model and simulation tool are built and their parameters are estimated. The model can then be used for predictions, either or not based on real time data. In all these steps errors are made: there are discrepancies between reality and observation, between observations and models, between models and simulations based on them. One step further, Huber et al. (1) introduce methods to assess the accuracy of traffic information.

A lot of research effort has been put into assessing the accuracy of the steps. Qualitative assessments are used to show whether certain characteristics or phenomena are well represented (2, 3, 4, 5). Yet, more effort has been put into quantification of accuracy and their application in model choice, parameter calibration, validation and sensitivity analysis (6, 7, 8, 9, 10, 11, 12, 13). However, the way accuracy is quantified is often straightforward and does not provide information on the type of error that was made. For example, variables are compared on a one-to-one basis: the predicted value of a selected variable (e.g. speed, headway, density) at certain locations and times is compared to the reference value at the same locations and times. For each location and time the difference is computed and the difference is averaged over the whole space and time domain. The averaging can be done in many ways (12), including the mean error (ME), mean absolute error (MAE) and root mean square error (RMSE). Therefore, the accuracy measure does not give insight into what causes the error and how improvements in the applied methodologies can help to increase the accuracy.

Our main contribution is the introduction of two new accuracy measures that do give insight into the type of error and take into account the specific features of traffic and pedestrian flow, such as chaotic behaviour and the importance of sharp transitions and the location and time when congestion occurs. The new accuracy measures apply the concepts of phase error and diffusion error and are introduced in the next section. They can be applied to both (one dimensional) road traffic flow and to (two dimensional) pedestrian flow, as we show in the Section ‘Case Studies’. We conclude by discussing the main results and future research directions.

ACCURACY MEASURES

Traffic and pedestrian flows show large and sometimes sudden changes in the values of important variables. Congestion occurs locally and there may be congestion with low speeds just upstream of an on ramp, while 500 meter downstream traffic is free flowing and speeds are high. Additionally, stop and go waves are characterised by sudden and steep changes in density and speed. Pedestrian flows also show complex phenomena such as lane formation and clogging upstream of an exit. We introduce accuracy measures that can handle these key phenomena and include high penalties if they are represented badly. We focus on three important features:

1. key phenomena with steep state changes over time and/or space such as a sudden drops in speed (e.g. at the tail of congestion) or sudden speed increases (e.g. just upstream of a bottleneck) (14, 15),
2. key phenomena with states alternating over space and possibly time such as stop and go waves in road traffic or lanes in pedestrian flows (16, 17, 18),
3. the speed or velocity of high density areas, causing spill back and possible blockage of ramps, exits or other parts of networks (2, 19, 20).

We select those features because they characterise traffic and pedestrian flows and the features distinguish them from many other types of flows such as most fluid flows.

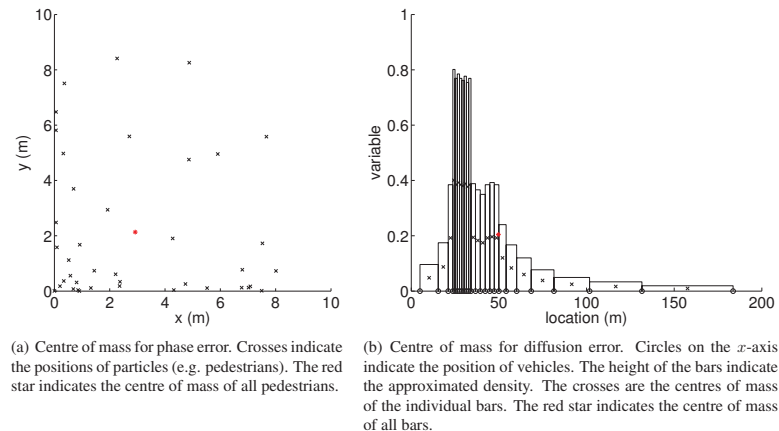


FIGURE 1 Examples of centre of mass used to define the phase error and diffusion error.

To measure the accuracy of a data set, a model, a simulation or a prediction, its variables (e.g. speed, velocities, densities or headways) are compared to those of a reference or ‘ground truth’. We propose two new accuracy measures for traffic and pedestrian flows. They are based on the concepts of phase error and diffusion error, as introduced in the next section. Unlike the traditional measures, the new measures take into account the specific features of traffic and pedestrian flows as described above. They do this by measuring whether the location of high density areas is reproduced correctly (phase error) and by measuring how well sharp transitions are reproduced (diffusion error).

Location of Centre of Mass and Phase Error

Both the phase error and the diffusion error use the concept of centre of mass to quantify the accuracy. In physics, the centre of mass is defined as the point at which the entire weight of a body may be considered as concentrated so that if supported at this point the body would remain in equilibrium in any position. We translate the location of the centre of mass \vec{X} as the point where all particles (vehicles or pedestrians) may be considered as concentrated. The concept is illustrated in Figure 1(a). We define the location of the centre of mass at time t . The location \vec{x}_n of the n -th particle is supposed to be known for all particles $n \in N$ present at time t . The location of the centre of mass $\vec{X}_{particle}$ is defined as follows:

$$\vec{X}_{particle} = \frac{1}{N} \sum_{n=1}^N \vec{x}_n \tag{1}$$

with $\vec{X}_{particle}$ and \vec{x}_n scalar in one dimensional road traffic flow and column vectors with two rows in two dimensional pedestrian flow.

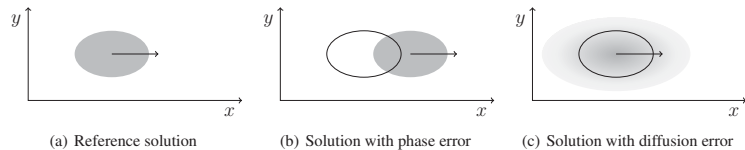


FIGURE 2 Three possible states of a crowd. The reference solution is considered as the ground truth, where all pedestrians are gathered within an ellipse shaped region (a). If a model predicts the location of the ellipse incorrectly, there is a phase error (b). If a model predicts that the pedestrians are more spread out over the region and the transition at the edge of the ellipse is too smooth, there is a diffusion error (c).

In a more generic case, location of the centre of mass within the domain Ω can be defined as:

$$\vec{X} = \frac{\int_{\vec{x} \in \Omega} \rho(\vec{x}) \vec{x} \, d\vec{x}}{\int_{\vec{x} \in \Omega} \rho(\vec{x}) \, d\vec{x}} \quad (2)$$

In a continuum model $\rho(\vec{x})$ would be the density at location \vec{x} , but it can also be interpreted as the probability that a particle is present at location \vec{x} . This definition is useful if the density at every point in the domain is known, or can be derived. If the domain can be subdivided into $i \in [1 : I]$ cells (road segments with a certain length or regions with a certain area), and the cells have constant density ρ_i , then (2) reduces to:

$$\vec{X}_{\text{continuum}} = \frac{\sum_{i=1}^I A_i \rho_i \vec{x}_i}{\sum_{i=1}^I A_i \rho_i} \quad (3)$$

with A_i the length (in one dimensional flow) or area (in two dimensional flow) of the i -th cell. Therefore, $A_i \rho_i$ is the (expected) number of particles within cell i . \vec{x}_i denotes the location of the centre of mass of the cell, which can be computed easily if the cell has a simple shape and ρ_i is constant within the cell. The definition in (3) is useful if data from loop detector is available or if continuum flow models are applied.

Phase Error

A phase error occurs if the average speed of the particles is too low or too high or if they move in the wrong direction, see Figure 2(b). This results in an inaccuracy in the location of patterns such as a high density/low speed region in a congestion pattern or in lanes formed in bi-directional pedestrian flow. The difference in the location of the centre of mass between the reference solution and the test solution is (here) a measure for the phase error:

$$\vec{E}_{\text{phase}} = \vec{X}_{\text{test}} - \vec{X}_{\text{ref}} \quad (4)$$

We note that in some cases a phase error is something to avoid, for example when one wants to predict if or when the tail of a queue upstream of a traffic light will reach the next upstream intersection. In other cases, such as with lanes formed in bi-directional pedestrian flow, a phase error

may only indicate that the high and low density/speed regions have ‘swapped’: e.g. pedestrians walk from left to right where they were supposed to walk from left to right and vice versa. This type of phase error does not pose a problem in most applications. The difference between two causes for phase errors implies that they always need an interpretation to understand whether the phase error should be avoided or it can be ignored.

Height of Centre of Mass and Diffusion Error

To quantify the diffusion error we use the concept of centre of mass in a slightly different way and use it to define the centre of mass in the direction of a fundamental variable such as density, speed or x - or y -velocity. The concept is illustrated in Figure 1(b). We define the centre of mass in the direction of the fundamental variable, as the centre of mass of the area under its profile. In the figure, the position of $N + 1$ vehicles is depicted, with x_n the location of the n -th particle. In this case, each bar indicates one vehicle. The centre of mass in the direction of variable f is the height of the combined centre of mass of all N bars:

$$F = \frac{1}{N} \sum_{n=1}^N \frac{f_n}{2} \quad (5)$$

We note that (5) also holds in two dimensional cases. Generalization of (5) without assuming each cell contains exactly one particle, but instead taking a continuum approach as in (3), leads to:

$$F_{\text{continuum}} = \frac{\sum_{i=1}^I A_i \rho_i f_i}{2 \sum_{i=1}^I A_i \rho_i} \quad (6)$$

with A_i the area of the i -th cell and ρ_i its density, which is assumed to be constant. Consequently, $A_i \rho_i$ is the number of particles in the i -th cell. Finally, further generalisation gives, similar to (2):

$$F = \frac{\int_{\vec{x} \in \Omega} \rho(\vec{x}) f(\vec{x}) \, d\vec{x}}{2 \int_{\vec{x} \in \Omega} \rho(\vec{x}) \, d\vec{x}} \quad (7)$$

Diffusion error

A diffusion error occurs if transitions between areas with a high value of the fundamental variable and with a low value become too smooth, see Figure 2(c). If the fundamental variable is density, high density regions get a lower density and vice versa. Therefore, if transitions are sharp, the centre of mass in density direction is higher than when transitions are smoother. The difference in the centre of mass in the direction of the fundamental variable between the reference solution and the test solution is a measure for the diffusion error:

$$E_{\text{diff}} = F_{\text{test}} - F_{\text{ref}} \quad (8)$$

CASE STUDIES

We discuss the concepts of phase and diffusion error further using some case studies. The case studies compare simulation results with analytical model solution (for road traffic, case 1), and with experimental data (for pedestrian flows, case 2). Errors are quantified using our newly developed accuracy measures and traditional ones. The results show that the new accuracy measures give insight into how accurate simulation results are and help to identify the cause of inaccuracies.

Case 1: Queue Upstream of Traffic Light

This case illustrates the application of the new accuracy measures to compare simulation results with the analytical model solution. We consider a hypothetical one lane road with a queue (density is jam density) of vehicles in front of a red light. Downstream of the queue the density is initially zero, the length of the queue is 2 kilometre and upstream of the queue the initial density is half of critical density. At time $t = 0$, the red light turns green and vehicles start driving. To determine the traffic state until the queue has solved, we use the LWR model (19, 21) with a parabolic-linear fundamental diagram (22) with maximum speed 33.3 m/s, critical speed 20.8 m/s, critical density 0.33 vehicles/m and jam density 0.2 vehicles/m.

Analytical and numerical solution

We solve this problem both analytically and by simulation. We use three different simulation methods: a minimum supply demand method with explicit time stepping (23, 24), an upwind method with explicit time stepping (25) and an upwind method with implicit time stepping (26). Furthermore, we vary the time step size to study the influence of the Courant-Friedrichs-Lewy (CFL)-number (27). With a small time step size, the CFL-number is small and a simulation takes long. However, with a too large time step size, the CFL-number is larger than 1 and the minimum supply demand method and the upwind method get unstable, resulting in unrealistic simulation results. The simulation results and the exact solution are shown in Figure 3 and 4.

Accuracy

Analysis of the numerical methods show that all methods are first order accurate, implying that the global error is proportional to the numerical resolution (28). However this does not mean that, in practice, the accuracy of the methods is similar. To measure the accuracy of the simulation results, they are compared to the exact solution. We do this only at time $t = 600$ s, which is representative for the whole time period. For all solutions we compute the centre of mass of the cross sections shown in Figure 4. The phase error and diffusion error are computed as in (4) and (8), respectively. For the diffusion error we use the density as variable: $f = \rho$. The errors are shown in Figure 5.

Results

The results in Figure 5 show that the minimum supply demand method always performs worse than the upwind explicit method. The performance of the upwind implicit method depends on the time step size: it gives relatively high accuracy if the CFL-number is small, while accuracy is low (errors are large) if it is large. For CFL-numbers larger than 1, the minimum supply demand method and the upwind explicit method become unstable and give very unrealistic results and they are therefore omitted from the figure. The accuracy of the upwind implicit method decreases with increasing CFL-number, but the accuracy is of the same order as that of the minimum supply demand method. Furthermore, the results show that, for the minimum supply demand and the upwind implicit method, the phase error depends more strongly on the CFL-number than the diffusion error. For the upwind explicit method the difference is not clearly visible from the graphs.

Discussion

These results give insight into when to use which numerical method. A large error, has to be weighted against a large computation time (i.e. small time steps). Furthermore, if the phase error is important in the application (for example if one wants to know whether congestion will pill

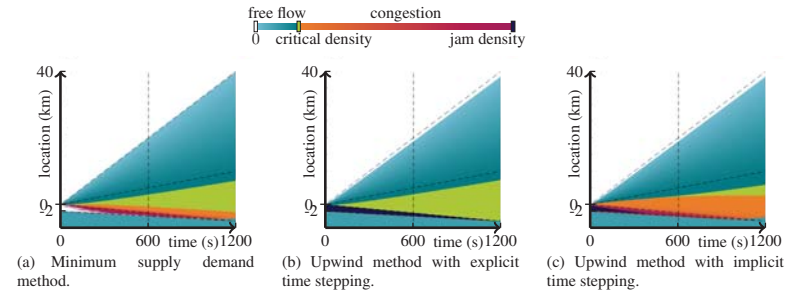


FIGURE 3 Densities computed with different numerical methods and fixed time step size. The vertical dashed line indicate the time for which cross sections are shown in Figure 4. The other dashed lines indicate sharp boundaries in the analytical solution.

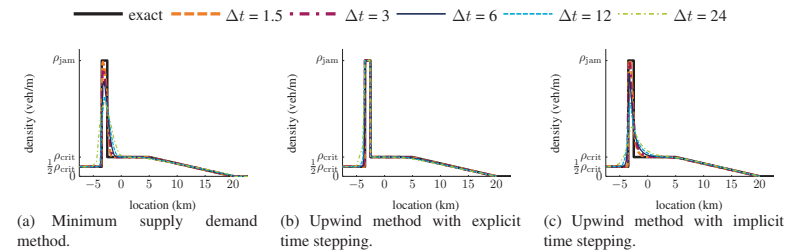


FIGURE 4 Density profiles at time $t = 600$ s with different numerical methods and time step sizes.

back until a certain off ramp), then also the CFL-number should be taken into account. If a small diffusion error is more important, than it is less important to chose a good CFL-number, as long as it is not larger than 1 with the minimum supply demand method or the upwind explicit method. We note that no final conclusion on an appropriate numerical method can be drawn from this very limited test case. Instead a wider range of tests such as in (26) should be performed. However, the test case shows how the new accuracy measures can be applied to support the choice of simulation method.

Case 2: Bidirectional Pedestrian Flow

In this case we illustrate how to compare data from an experiment with bidirectional pedestrian flows (29) with simulation results and how to apply the results in parameter estimation. We consider a 10 meter long and 4 meter wide corridor. Two groups of pedestrians cross the corridor in opposite directions, thus creating a bi-directional flow.

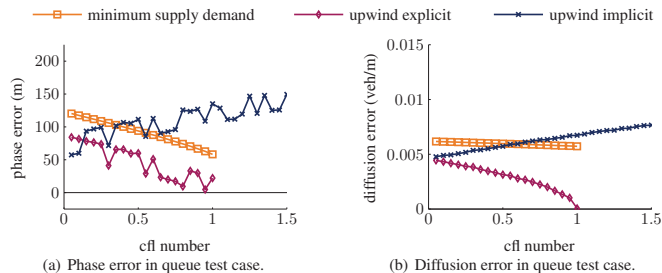


FIGURE 5 Accuracy of solutions at time $t = 600$ s with different numerical methods and time step sizes/CFL numbers.

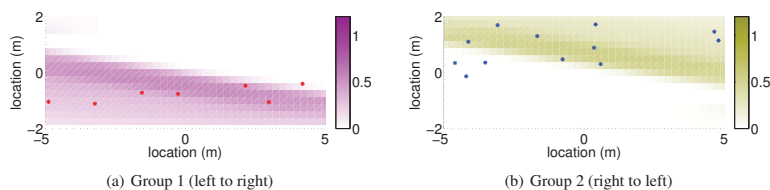


FIGURE 6 Typical class specific densities in simulation (coloured areas, purple and yellow) and positions of pedestrians in preprocessed experimental data (red and purple stars), both at time $t = 350$ s.

Experimental data

The data that was gathered from this experiment and is used here, includes positions in both x - and y -direction every 0.1 seconds, for 7 minutes. The data also includes identification numbers to trace individual pedestrians. We preprocessed the data to extract velocities of individual pedestrians every 0.1 seconds and to assign each pedestrian to a group: group 1 when entering on the left and leaving on the right, group 2 for pedestrians walking in the opposite direction. The preprocessed data shows that the setup resulted in lane formation. Most frequently, there are 2 lanes with pedestrians walking on their right hand side of the corridor. In some situations, there are 1, 3 or 4 lanes (30). A typical situation is shown in Figure 6.

Simulation results

The experimental data is compared to simulation results. These were obtained using a continuum pedestrian flow model (31) and simulation method (32). The simulation was performed for a

TABLE 1 Settings and some results of pedestrian simulation.

parameter set	β_u	β_o	N	resulting lanes
1	0.80	2.30	705	right
2	0.70	1.36	708	left
3	0.63	0.63	707	none
experimental data	na	na	709	right

corridor of the same width (4 meters) but longer (50 meters). Only the results for the centre 10 meter were used for comparison with the experimental data, see Figure 6.

In the simulation, pedestrians enter from both sides, with inflow demand equal to capacity. Group 1 moves from the left to the right, group 2 moves from right to left. Parameters were chosen to have realistic values: free flow speed is 1.34 m/s and jam density is 5.4 pedestrians/m². However, to compare the accuracy for different simulation results we used three different sets of parameter values for the avoidance of the own group (β_u) and the avoidance of the other group (β_o). For more details about the model and its parameters we refer to (31). We chose the parameter values such that the total number of pedestrians crossing the area would be very similar to those in the experimental data with $N = 709$. Furthermore, the parameters were chosen to give clearly distinguishable results, with either two lanes with pedestrians walking on their right hand side of the corridor (set 1) or on their left hand side (set 2), or with no lanes at all (set 3). We did not consider parameter settings resulting in more than two lanes. The settings are summarised in Table 1.

Accuracy

For both the experimental data and the simulation results, the centres of mass are computed and compared. We compute the phase error as in (4), but for both groups separately. For the diffusion error, we apply (8), again for both groups separately. We use the centre of mass of the reference solution F_{ref} as in (5) with $f_n = v_{x,n,u}$ the velocity in x -direction of the n -th pedestrian of group u and with N the total number of pedestrians of group u . Furthermore, the centre of mass of the test solution F_{test} is as in (6) with $f_i = v_{x,i,u}$ the velocity in x -direction of group u in the i -th cell. Also for the density ρ_i only pedestrians of group u are taken into account, i.e. ρ_i is the number of pedestrians of group u in cell i . We use the centre of mass in x -velocity direction instead of in density (ρ) direction because the density is not straightforward to determine from the experimental data, while the x -velocity is relatively simple to extract from the position data. Furthermore, it gives better insight into the differences in the results than the speed or y -velocity, because speed is always positive and in the experimental data close to free flow speed of the experiment (1.34 m/s) and because y -velocity is mostly close to zero.

We also compared the parameter sets with respect to traditional error measures. To compute the errors, at each time step, the experimental data for each pedestrian was compared with the simulations results at that location. The RMSE of the speed of each group at time t is defined as:

$$E_{\text{RMSE},|v_u|}(t) = \frac{\sqrt{\sum_{i \in I} [|v_{u,\text{ref}}(x_i, y_i, t)| - |v_{u,\text{test}}(x_i, y_i, t)|]^2}}{N_{\text{test}}(t)} \quad (9)$$

with (x_i, y_i, t) the position of the i -th pedestrian in the experimental data at time t , $u \in \{1, 2\}$ the group, $|v_{u,\text{ref}}(x_i, y_i, t)|$ its speed, $|v_{u,\text{test}}(x_i, y_i, t)|$ the speed at this location and time according

to the simulation and $N_{\text{test}}(t)$ the number of pedestrians in the area at time t according to the experimental data. The ME of the speed of each group at time t is defined as:

$$E_{\text{ME},|v_u|}(t) = \frac{\sum_{i \in I} [|v_{u,\text{ref}}(x_i, y_i, t)| - |v_{u,\text{test}}(x_i, y_i, t)|]}{N_{\text{test}}(t)} \quad (10)$$

The RMSE of the velocity in x -direction is defined as:

$$E_{\text{RMSE},v_x}(t) = \frac{\sqrt{\sum_{i \in I} [v_{x,\text{ref}}(x_i, y_i, t) - v_{x,\text{test}}(x_i, y_i, t)]^2}}{N_{\text{test}}(t)} \quad (11)$$

with the velocity in x -direction the weighted average velocity of both groups in the simulation:

$$v_{x,\text{test}}(x_i, y_i, t) = \frac{\rho_1 v_{x,1} + \rho_2 v_{x,2}}{\rho_1 + \rho_2} \quad (12)$$

The MAE of the velocity in x -direction is defined as:

$$E_{\text{MAE},v_x}(t) = \frac{\sum_{i \in I} |v_{x,\text{ref}}(x_i, y_i, t) - v_{x,\text{test}}(x_i, y_i, t)|}{N_{\text{test}}(t)} \quad (13)$$

Results

The errors are shown in Figures 7–9 and Tables 2 and 3. The figures show the development of the errors over time, the table shows the average error over the last 200 seconds. We note that the values of the errors vary largely over time, unlike in the road traffic test case. This shows that we would not be able to draw reasonable conclusions if we would only study the error at one moment in time as we did in the previous case.

TABLE 2 Errors in the pedestrian simulation, traditional measures

parameter set	$\vec{E}_{\text{RMSE}, v_1 }$	$\vec{E}_{\text{RMSE}, v_2 }$	$\vec{E}_{\text{RMSE},v_x}$	$\vec{E}_{\text{ME}, v_1 }$	$\vec{E}_{\text{ME}, v_2 }$	\vec{E}_{MAE,v_x}
1	0.066	0.070	0.31	-0.028	-0.046	0.73
2	0.065	0.069	0.54	-0.030	-0.049	1.86
3	0.065	0.068	0.34	-0.027	-0.046	1.27

TABLE 3 Errors in the pedestrian simulation, new measures

parameter set	E_{phase, x_1}	E_{phase, x_2}	E_{phase, y_1}	E_{phase, y_2}	E_{diff, v_1}	E_{diff, v_2}
1	-0.57	0.64	-0.42	0.19	0.033	0.049
2	-0.68	0.51	1.27	-1.25	0.032	0.044
3	-0.45	0.52	0.41	-0.64	0.026	0.041

Most traditional error measures do not discriminate between the parameter sets: the RMSE of the group specific speeds ($|v_1|$ and $|v_2|$) are almost the same for each parameter set. The same holds for the ME of the group specific speeds. The only difference can be found in the velocities in x -direction, both the RMSE and the MAE show different values for different parameter sets. However, the error is largest if the lanes are swapped (i.e. pedestrians walk on their left hand side

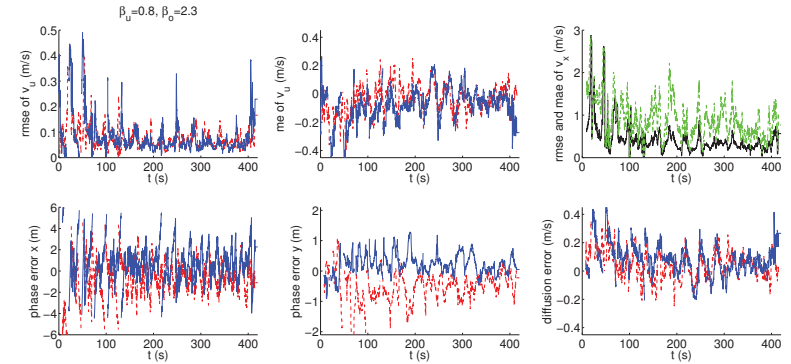


FIGURE 7 Errors using parameter set 1. Red dashed: group 1 (walking from left to right), blue solid: group 2 (walking from right to left), black solid: RMSE of v_x , green dashed: MAE of v_x .

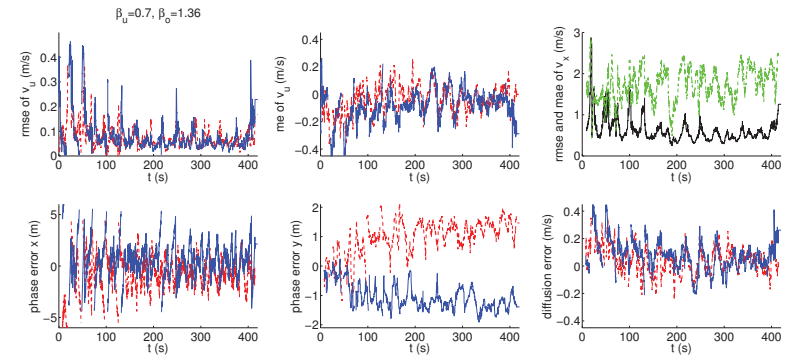


FIGURE 8 Errors using parameter set 2. Red dashed: group 1 (walking from left to right), blue solid: group 2 (walking from right to left), black solid: RMSE of v_x , green dashed: MAE of v_x .

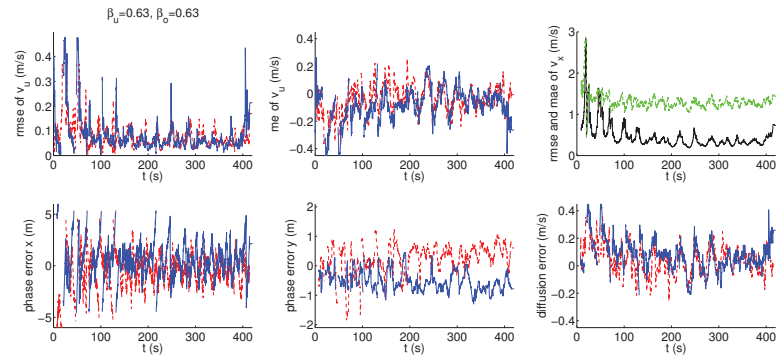


FIGURE 9 Errors using parameter set 3. Red dashed: group 1 (walking from left to right), blue solid: group 2 (walking from right to left), black solid: RMSE of v_x , green dashed: MAE of v_x .

of the corridor, instead of at their right hand side). The difference in RMSE of v_x between set 1 (correct lane formation) and set 3 (no lanes) is small.

The phase errors for both groups in x -direction are similar for each parameter set and we do not see any pattern. However, the other new measures do discriminate between the parameter sets. The absolute phase error (pe in y_1 and in y_2) is largest for parameter set 2, as expected. However, this is combined with a diffusion error of similar size than diffusion errors with the other parameter sets. The diffusion error is small, but clearly largest for set 1 and smallest for set 3.

Discussion

We note that in all simulations, the total number of pedestrians crossing the corridor is (almost) the same as in the experimental data. Therefore, an error measure based on the total flux (such as used by Kretz et al. (9)) would not be able to discriminate between the parameter sets. An other traditional error measure is based on the likelihood of the simulation result (11). This may be an appropriate approach to estimate parameters. However, it does not give any insight into the type of inaccuracies and how they may be reduced.

The results of the traditional accuracy measures give little insight in the cause of the error and would prefer parameter set 3 over set 2, while in most applications, the lane formation is essential and parameter set 2 would yield better results than set 3. The results for the phase error in y -direction are similar to those for the RMSE and MAE of v_x . However, the phase error gives insight into the type of error: the location of the lanes is predicted wrongly in set 2. The diffusion error does not add much insight because the speeds both in the experimental data and in the simulation are always close to the free flow speed of 1.34 m/s. Furthermore, in the experiments, three or four lanes were formed in some instances, while in the simulations there are only two. For a test case with more variation in the speeds or with a simulation setting with more lanes, the diffusion error may be able to discriminate between different parameter sets.

DISCUSSION, CONCLUSION AND OUTLOOK

We have introduced two new accuracy measures, phase error and diffusion error. They give better insight in the type of error and take into account specific features of traffic and pedestrian flows. The accuracy measures quantify the error using a method based on the centre of mass of a region with high density or speed. The new accuracy measures are more meaningful in traffic and pedestrian flow applications than traditional methods. Because of their generic formulation, they can be used in a wide range of applications: from the assessment of a data collection method to the quantification of the accuracy of traffic information.

We applied the new accuracy measures to two test cases. In the first case we compared different simulation methods for road traffic flow (one-dimensional), in the second case we compared different parameter settings of a pedestrian flow model. Both cases show that the way the accuracy is defined is important. A good result according to one accuracy measure, may be a bad result according to another accuracy measure. Furthermore, some accuracy measures fail to discriminate between the different test settings and give similar accuracies independent of the simulation method or parameter set. The new accuracy measures give insight into the type of error by showing, for example, that the phase error is small, while the diffusion error is large. This indicates that, if one wants to increase the accuracy, the model or simulation method should be improved to reduce diffusion.

It is important to note that, especially for the phase error, the interpretation of the error remains important. In some cases (such as while predicting the spill back of a queue until the upstream intersection), a large phase error should be avoided, while in other cases (such as lane formation in bi-directional pedestrian flow) a phase error may only indicate that the lanes are swapped.

The current applications focus on simple problems and networks with different phenomena at different location have not yet been considered. The new accuracy measures could be used as building blocks for an assessment method for accuracy on networks. This can for example be done by splitting the networks into parts and isolating areas with different layout or flow characteristics. Furthermore, we have now explored the applications for one-dimensional road traffic flow and for two-dimensional pedestrian flow. However, the methods may be extended to other applications such as network flows using a Network Fundamental Diagram (33).

Another future research direction is to extend the measures to changes over time. In our current applications, either the error at just one moment was considered, or the average error over a longer period of time. Time could be considered explicitly in the measures by also comparing the centre of mass in time direction. This may especially improve the quantification of the phase error.

Finally, the measures may be extended for heterogeneous flow including for example trucks and passenger cars or different types of pedestrians. In (26) we have developed a method to give trucks a higher weight when computing the centre of mass than passenger cars. This method can be extended to pedestrian flows.

ACKNOWLEDGEMENT

This research is performed as part of the NWO Aspasia grant of Daamen (first and second author) and as part of the NWO-VICI project of Hoogendoorn (third author).

REFERENCES

- [1] Huber, G., K. Bogenberger, and R. L. Bertini, New Methods for Quality Assessment of Real Time Traffic Information. In *Transportation Research Board 93th Annual Meeting Compendium of Papers*, Washington D.C., 2014.
- [2] Ward, J. A. and R. E. Wilson, Criteria for convective versus absolute string instability in car-following models. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 2011.
- [3] del Castillo, J., Three new models for the flow-density relationship: Derivation and testing for freeway and urban data. *Transportmetrica*, Vol. 8, No. 6, 2012, pp. 443–465.
- [4] van Wageningen-Kessels, F. L. M., B. van 't Hof, S. P. Hoogendoorn, J. W. C. van Lint, and C. Vuijk, Anisotropy in generic multi-class traffic flow models. *Transportmetrica A: Transport Science*, Vol. 9, No. 5, 2013, pp. 451–472.
- [5] Duives, D. C., W. Daamen, and S. P. Hoogendoorn, State-of-the-art crowd motion simulation models. *Transportation Research Part C: Emerging Technologies*, Vol. 37, No. 0, 2013, pp. 193–209.
- [6] Punzo, V. and F. Simonelli, Analysis and Comparison of Microscopic Traffic Flow Models with Real Traffic Microscopic Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1934, 2005, pp. 53–63.
- [7] Hoogendoorn, S. P. and W. Daamen, Microscopic Calibration and Validation of Pedestrian Models: Cross-Comparison of Models Using Experimental Data. In *Traffic and Granular Flow '05* (A. Schadschneider, T. Pöschel, R. Kühne, M. Schreckenberg, and D. Wolf, eds.), Springer Berlin Heidelberg, 2007, pp. 329–340.
- [8] Ossen, S., S. P. Hoogendoorn, and B. G. H. Gorte, Interdriver Differences in Car-Following: A Vehicle Trajectory-Based Study. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1965, 2006, pp. 121–129.
- [9] Kretz, T., S. Hengst, and P. Vortisch, Pedestrian Flow at Bottlenecks - Validation and Calibration of Vissim's Social Force Model of Pedestrian Traffic and its Empirical Foundations. In *3rd International Symposium on Transport Simulation (IST08): symposium proceedings*, 2008.
- [10] Kesting, A. and M. Treiber, Calibrating Car-Following Models by Using Trajectory Data: Methodological Study. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2088, 2008, pp. 148–156.
- [11] Robin, T., G. Antonini, M. Bierlaire, and J. Cruz, Specification, estimation and validation of a pedestrian walking behavior model. *Transportation Research Part B: Methodological*, Vol. 43, No. 1, 2009, pp. 36–56.
- [12] Ciuffo, B., V. Punzo, and M. Montanino, *The Calibration of Traffic Simulation Models: Report on the assessment of different Goodness of Fit measures and Optimization Algorithms*. Joint Research Centre - European Commission, 2012.

- [13] Ciuffo, B., J. Casas, M. Montanino, J. Perarnau, and V. Punzo, Gaussian Process Metamodels for Sensitivity Analysis of Traffic Simulation Models. *Transportation Research Records: Journal of the Transportation Research Board*, Vol. 2390, 2013, pp. 87–98.
- [14] Daganzo, C. F., Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B: Methodological*, Vol. 29, No. 4, 1995, pp. 277–286.
- [15] Chalons, C., P. Goatin, and N. Seguin, General constrained conservation laws. Application to pedestrian flow modeling. *Networks and Heterogeneous Media*, Vol. 8, No. 2, 2013, pp. 433–463.
- [16] Zielke, B. A., R. L. Bertini, and M. Treiber, Empirical Measurement of Freeway Oscillation Characteristics: An International Comparison. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2088, 2008, pp. 57–67.
- [17] Johansson, A., D. Helbing, H. Z. Al-Abideen, and S. Al-Bosta, From Crowd Dynamics to Crowd Safety: A Video-Based Analysis. *Advances in Complex Systems*, Vol. 25, 2008, pp. 9–41.
- [18] Jiang, Y., T. Xiong, S. Wong, C.-W. Shu, M. Zhang, P. Zhang, and W. H. Lam, A reactive dynamic continuum user equilibrium model for bi-directional pedestrian flows. *Acta Mathematica Scientia*, Vol. 29, No. 6, 2009, pp. 1541 – 1555.
- [19] Lighthill, M. J. and G. B. Whitham, On Kinematic Waves II: A Theory of Traffic Flow on Long Crowded Roads. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 229, No. 1178, 1955, pp. 317–345.
- [20] Colombo, R. M., P. Goatin, and M. D. Rosini, A macroscopic model for pedestrian flows in panic situations. In *4th Polish-Japan Days*, Madralin, Poland, 2010, Vol. 32, pp. 255–272.
- [21] Richards, P. I., Shock Waves on the Highway. *Operations Research*, Vol. 4, No. 1, 1956, pp. 42–51.
- [22] Smulders, S., Control of freeway traffic flow by variable speed signs. *Transportation Research Part B: Methodological*, Vol. 24, No. 2, 1990, pp. 111–132.
- [23] Daganzo, C. F., The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, Vol. 28, No. 4, 1994, pp. 269–287.
- [24] Lebacque, J.-P., The Godunov scheme and what it means for first order traffic flow models. In *Transportation and Traffic Theory: Proceedings of the 13th International Symposium on Transportation and Traffic Theory, 1996* (J.-B. Lesort, ed.), Pergamon, 1996, pp. 647–677.
- [25] Leclercq, L., J. Laval, and E. Chevallier, The Lagrangian coordinates and what it means for first order traffic flow models. In *Transportation and Traffic Theory 2007* (R. E. Allsop, M. G. H. Bell, and B. G. Heydecker, eds.), Elsevier, Oxford, 2007, pp. 735–753.

- [26] van Wageningen-Kessels, F. L. M., *Multi class continuum traffic flow models: Analysis and simulation methods*. Ph.D. thesis, Delft University of Technology/TRAIL Research school, Delft, 2013.
- [27] Courant, R., K. Friedrichs, and H. Lewy, On the partial difference equations of mathematical physics. *IBM Journal*, 1967, pp. 215–234.
- [28] LeVeque, R. J., *Finite volume methods for hyperbolic problems*. Cambridge texts in applied mathematics, Cambridge University Press, Cambridge, 2002.
- [29] Daamen, W. and S. P. Hoogendoorn, Controlled experiments to derive walking behaviour. *Journal of Transport and Infrastructure Research*, Vol. 3, No. 1, 2003, pp. 39–59.
- [30] Campanella, M. and W. D. S. P. Hoogendoorn, Calibration of pedestrian models with respect to lane formation self-organisation: Exploring chaos to obtain coherent behaviour. In *TRAIL Congress (DVD)*, Rotterdam, 2008.
- [31] Hoogendoorn, S. P., F. L. M. van Wageningen-Kessels, W. Daamen, and D. C. Duives, Continuum Modelling of Pedestrian Flows: From Microscopic Principles to Self-Organised Macroscopic Phenomena. *Physica A*, in press.
- [32] van Wageningen-Kessels, F. L. M., W. Daamen, and S. P. Hoogendoorn, The two-dimensional Godunov scheme and what it means for macroscopic pedestrian flow models, 2015, extended abstract accepted, full paper in preparation for submission to ISTTT and subsequent publication in Transportation Research series.
- [33] Daganzo, C. F. and N. Geroliminis, An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transportation Research Part B: Methodological*, Vol. 42, No. 9, 2008, pp. 771–781.

THE EFFECT OF STOCHASTIC VOLATILITY IN PREDICTING HIGHWAY BREAKDOWNS

Eric M. Laflamme^{a,*} and Paul J. Ossenbruggen^b

^aDepartment of Mathematics, Plymouth State University, Plymouth, NH, emlaflamme@plymouth.edu, 603-724-5336

^bDepartment of Mathematics and Statistics, University of Massachusetts Boston, Boston, MA, pjo@unh.edu,

603-531-3330

*Corresponding author

ABSTRACT

In this paper we pursue the effect of volatility on the probability of highway breakdown. Because daily aggregated flow values exhibit dissimilar levels of variability throughout the day, a stochastic volatility (SV) model was pursued. Under this format, assuming traffic flow volatility follows an autoregressive process of order one, volatility was estimated at each 15-minute time step per day during the collection period of 205 days. A computational Bayesian format was used to fit parameters of 205 SV models to data collected from one bottleneck site along Interstate 93 in Salem, NH. Fitted results show that volatility estimates successfully capture flow variability observed from the traffic data. To evaluate the effect of SV on breakdown, a sampling scheme was created in which sampled demands and capacities were compared. To estimate demand, a signal was first extracted from the flow aggregates using a Functional Data Analysis (FDA) model and then combined with a sampled SV estimate. For stochastic capacity, daily flow maxima were used as either censored or uncensored estimates of capacity based on breakdown occurrence. By extreme value theory, these capacity values are distributed as a Generalized Extreme Value (GEV) distribution. From a GEV model fitted to our data, capacity estimates are sampled and compared to SV-based demand to produce breakdown probabilities by time of day. Finally, our breakdown probabilities are compared to those estimated by the sample data only to illustrate the effect of SV on breakdown.

THE EFFECT OF STOCHASTIC VOLATILITY IN PREDICTING HIGHWAY BREAKDOWNS

Eric M. Laflamme^a and Paul J. Ossenbruggen^b

^aDepartment of Mathematics, Plymouth State University, Plymouth, NH, emlaflamme@plymouth.edu

^bDepartment of Mathematics and Statistics, University of Massachusetts Boston, Boston, MA, pjo@unh.edu

INTRODUCTION

Theoretically, highway breakdown will occur when flow q_t , some flow (q) at time (t), equals or exceeds roadway capacity c . However, because breakdowns do not necessarily occur at the same demand levels (Elefteriadou et al., 1995), our breakdown protocol $q_t \geq c$ cannot be confirmed with field data (Brilon et al., 2007; Lorenz and Elefteriadou, 2001). In light of this, breakdown will be treated as a chance event with the following probability model:

$$P(D_t \geq C)$$

where travel demand at time t , D_t , and roadway capacity, C , are both considered random variables. In order to estimate this probability expression, we must identify suitable models for both traffic demand and capacity. A complication arises as flow values, values used to calibrate our models, are typically very noisy. To address this, to mitigate the irregular fluctuations, flow values are typically extracted from coarse aggregates. That is, traffic flow variability (noise) is removed to identify a realistic signal. What if we could estimate the natural variability in demand and re-introduce it into our breakdown prediction? In so doing, how would this volatility in demand affect the performance of the roadway? Will traffic noise increase the chance of breakdown? In this work we aim to answer these questions and quantify the effect of flow volatility on the probable occurrence of breakdown.

APPROACH

A dominant characteristic of freeway traffic flow is the time-varying volatility. That is, a roadway commonly experiences distinct periods of contrasting levels of flow variability (Figure 1). Such behavior is also commonly observed in financial or economic time series. To account for this type of behavior, Taylor (1982) suggests that volatilities evolve in a stochastic nature, an approach commonly referred to as the ‘stochastic volatility’ (SV) model. So, in our expression for the probability of breakdown, $P(D_t \geq C)$, for any day,

$$D_t \sim N(q_t, \sigma_t),$$

or demand at time t is normally distributed about some flow q_t with a time-indexed standard deviation (Note the distinction between demand and flow, where, going forward, demand is a combination of known flow and random variability). The unique feature of the SV model is the time-varying standard deviation, σ_t , which is itself introduced as a random variable.

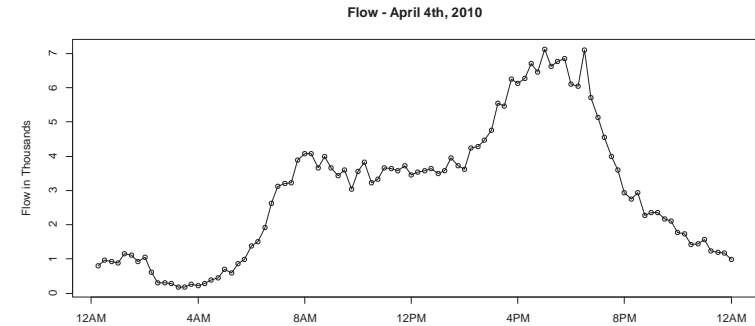


FIGURE 1 Illustration of time-varying volatility in traffic flows for a typical weekday during the collection period. We observe that flow values are more volatile during morning and evening commuting hours.

DATA

Data was collected by the State of New Hampshire (NH) Department of Transportation along northbound lanes of interstate I-93 at one location in Salem, NH that observes both daily breakdowns and heavy demand. Data collection was made possible by a side-fire radar unit located just north of an exit ramp, exit 1, and just south of an entrance ramp to I-93. Just north of the radar device (downstream), I-93 is physically constricted from three to two lanes. The positioning of this device is ideal as data collected here fit perfectly to the theoretical framework of flow-based capacity analyses (Minderhoud et al., 1997). As stated by Brilon et al. (2005), the capacity of the freeway segment is analyzed most precisely at or slightly upstream of a bottleneck.

The radar device was scheduled to intermittently measure traffic at irregular but frequent time periods about 1 minute apart. The data collection period occurred between April 1 and November 30, 2010, a total of 244 days. The observations (raw data) consist of the following measurements: vehicle counts, average speed, occupancy, and speed of individual vehicles observed over the interval. During this collection period, there were several scheduled days where the radar devices were shut down for maintenance. In the end, after omitting these days where no traffic measurements were recorded, a total of 205 days were retained for our analysis.

Aggregation

Because radar data are collected over very short, irregular time intervals, these measurements were aggregated into uniform intervals. The HCM recommends that transportation engineers use aggregates (intervals) no shorter than 15-minutes in order to ensure ‘stable’ traffic flow rates. Longer intervals are especially suitable for macroscopic/speed-flow analyses and intervals shorter than 5-minutes should be avoided (Highway Capacity Manual,

2000). That is, for shorter intervals, it is possible to observe speeds and flows that simply cannot be sustained over longer periods, and these measurements are not accurate representations of traffic conditions. As stated by Greenwood et al. (2007), analyzing capacity is a macroscopic endeavor related to management, policy, design and regional speed/flow comparisons, thus stable, longer interval durations are appropriate. While exact lengths vary among analyses, 15-minute intervals are commonly used for capacity studies (See, for example, Agyemang-Duah and Hall, 1991; Yang and Zhang, 2005; Lorenz and Elefteriadou, 2001; Elefteriadou and Lertworawanich, 2002; Minderhoud et al., 1997). For these reasons, and to produce results that are directly comparable to those put forth by the HCM, observations were binned into 15-minute aggregates.

From each of the raw vehicle counts (total volume, or number of vehicles, observed during a precisely defined interval), the corresponding interval length during which they were observed and a flow rate, or equivalent hourly rate of vehicles passing a point (the 'traffic flow rate,' by definition), was simply calculated. This then yields flow rates that are expressed in the customary scale of vehicles per hour (vph), but based on short, unstable time intervals. Next, these flow rates were binned into non-overlapping, sequential 15-minute intervals based on original observation times. Lastly, for each bin, the component flow rates were averaged to produce aggregated flow (q) rates for each 15-minute interval during the collection period. Thus, we have 96 estimates per day, or $t = 1, \dots, 96$. Addressing traffic speed (u) is a simpler task as precise speed (in mph) and time measurements are recorded for every vehicle observed during the collection period. Speed aggregates are then calculated by first separating all observations into 15-minute bins, and then simply taking their average.

CAPACITY DATA

Daily Maxima as Estimates of Capacity

Modern methods of estimating breakdown probabilities are based on stochastic capacity, or treating capacity as a random variable (See, for example, Brilon et al., 2005; Lorenz and Elefteriadou, 2001). That said, no universally accepted measure of capacity has been established. Breakdown flows, however, those flows measured immediately before the onset of congestion, have been widely adopted as estimators of capacity (See, for example, Elefteriadou and Lertworawanich, 2002; Brilon et al., 2005; Lorenz and Elefteriadou, 2001; Minderhoud et al., 1996). However, one concern when using breakdown flows is that higher flows or daily flow maxima often occur prior to congested conditions. These cases suggest that breakdown flows are actually underestimating capacity since higher flows were observed just prior to breakdown. By instead considering *daily flow maxima*, we would capture these higher flows prior to breakdown and ultimately obtain more representative measures of capacity, the maximum flow that a roadway can sustain. Another concern when using breakdown flows is that they are highly dependent on subjective, and often arbitrary, breakdown identification criteria. Daily flow maxima, of course, have the advantage of being independent of the methods used to identify breakdowns.

Another commonly used estimate of freeway capacity is the maximum pre-breakdown flow (See, for example, Hall and Agyemang-Duah, 1991; Hall et al., 1992), or the maximum sustained flow prior to breakdown. Based on aggregated traffic data, Elefteriadou and Lertworawanich (2002) provide evidence that distributions of breakdown flows and maximum

pre-breakdown flows are statistically similar and, by extension, that maximum pre-breakdown flows are suitable estimates for capacity. Because daily flow maxima typically occur prior to breakdowns, in practice, daily flow maxima closely resemble maximum pre-breakdown flows, although the two are conceptually different. Thus, using daily flow maxima to estimate capacity has many of the same benefits as using pre-breakdown flows. For one, both maxima and pre-breakdown flows account for cases (as described above) where flows decrease prior to congestion, a so-called 'lingering' or lagged effect, and are better representations of the true maximum throughput of the roadway. However, while maxima and pre-breakdown flows are somewhat similar, maxima have the advantage of being consistently available. That is, extraction of pre-breakdown flows is completely dependent on breakdowns, while daily maxima, on the other hand, have no such dependence and may be obtained on a regular basis regardless of breakdown occurrence. Using maxima, then, is especially beneficial when breakdowns are rare as such cases would yield relatively few pre-breakdown flows. Laflamme (2013) presents a detailed discussion on the use and advantage of using daily maxima as capacity values.

Censoring

As stated by the HCM, the capacity for a given facility is the flow rate that can be achieved repeatedly for peak periods of sufficient demand (Highway Capacity Manual, 2000). Therefore, not every maximum flow is a suitable estimate of freeway capacity, and only those daily maxima associated with 'sufficiently high demand', demand typically resulting in breakdown, should be considered as such. On days where demand is insufficient, when breakdowns are not observed, daily conditions are not adequate (sufficiently extreme) to assess the true capacity of the roadway. In these cases, under a survival analysis premise, the corresponding maxima are deemed censored (right-censored) capacity values as the roadway can surely service higher demands. That is, breakdowns would occur at some higher flow rates, and the resulting capacity, the maximum daily flows, would necessarily be larger than the observed value. Alternatively, we may simply consider these cases as incomplete data records where the true maximum, the capacity, is simply missing. Despite the incompleteness associated with censored values, they still contain valuable information and will therefore be considered in the calibration of our capacity distribution (See, for example, Geistefeldt, 2010). In a comparison of capacity distributions approaches, Geistefeldt and Brilon (2009) found that using censored data achieves significantly more precise estimates, especially at higher quantiles. Lastly, our censoring designation establishes a correspondence between extreme flows and breakdown which we intuitively know exists.

The Generalized Extreme Value Distribution

Our capacity data, daily maxima extracted from the traffic stream measurements, are approximated by the Generalized Extreme Value distribution (GEV). That is, $C \sim GEV(\alpha, \beta, \xi)$, where C represents the capacity random variable. The GEV is three-parameter distribution given by the following form:

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\alpha}{\beta}\right)\right]^{-1/\xi}\right\}$$

defined on $\{x: 1 + \frac{\xi(x-\alpha)}{\beta} > 0\}$ with α and β the respective location and scale parameters. The shape parameter of the GEV, ξ , characterizes the rate of tail decay, where $\xi > 0$, $\xi = 0$, and $\xi < 0$ correspond to data with heavy tails, light tails, and short tails, respectively. For our data, in the expression above, we replace the random variable X with capacity data C as both censored and uncensored (observed) daily flow maxima. By theory, our capacity data, the maximal flows, can be suitably and accurately approximated by the GEV distribution.

Despite violating the model assumption, that a series of independent random variables define the block maxima, the conclusion that the block maxima have a GEV distribution is still reasonable (Coles, 2001). Quantile plots for daily maxima show strong correspondence between observed and predicted values, which illustrates the appropriateness of the model class and justifies the use of the GEV for our data. Additionally, the GEV approach assumes daily maxima be identically distributed which, from observing the 8 months of data, appears to be the case. That is, we observe no true seasonality ('heavy' season), no trend, no oscillations nor systematic patterns in maximal values.

TRAVEL DEMAND

When forecasting $P(D_t \geq C)$, simply specifying demand as a constant flow value, q_t , would naively suggest that the travel demand is known with certainty, and breakdown prediction would be $P(q_t \geq C)$. The noise observed in the daily demand profile, as seen in Figure 1, would not play a role in forecasting breakdown. We simply do not know that this is the case. In our approach, we consider $D_t = q_t + N(0, \sigma_t)$ for each of the 205 days. To investigate whether or not noise in flow has a destabilizing effect on reliability, we will assess the effect that both q_t and σ_t have on our breakdown probability estimates.

Stochastic Volatility

For each of the 205 days and at each of the 96 time periods, stochastic volatility was estimated. Following Kastner and Fruhwirth-Schnatter (2013), the variance in flow measurements (q_t), σ_t , for each day is assumed to have the following properties:

$$\begin{aligned} \sigma_t | h_t &\sim N(0, \exp h_t) \\ h_t | h_{t-1}, \mu, \phi, \lambda_\eta &\sim N(\mu + \phi(h_{t-1} - \mu), \lambda_\eta^2) \\ h_0 | \mu, \phi, \lambda_\eta &\sim N(\mu, \lambda_\eta^2 / (1 - \phi^2)) \end{aligned}$$

where $\exp(h_t)$ is called the latent, time-varying volatility that follows a stochastic evolution. Since h_t above depends on h_{t-1} , σ_t is a one-step ahead forecast that depends on σ_{t-1} . Moreover, σ_t has a variance that is not allowed to vary unrestrictedly with time, but rather is an autoregressive process of order one. We note that this model was fitted to each of the 205 days independently, thus allowing the parameters μ , ϕ , and λ_η to vary by day. This approach was deemed necessary in order to capture the day-to-day variation observed in the traffic record.

FUNCTIONAL DATA ANALYSIS (FDA) MODELING

An integral tool used to analyze traffic breakdown is functional data analysis (FDA), or functional data models. FDA models are used to extract the signal from the 15-minute aggregates which, although they are accurate representations of speed and flow, are still very 'noisy.' Furthermore, irregularities in the data collection process, or missing measurements, may result in aggregates calculated from few observations. Because these raw traffic observations can be erratic, because short intervals may identify traffic estimates that can only be sustained over the short term, aggregates based on few of these raw observations may be similarly noisy. To address this issue, functional data analysis (FDA) is employed to produce summaries, or smooth representations, of the data (both speed and flow) over a fine time scale, to extract the signal from the noise. These resulting interpolating smooths will be herein referred to as 'traces' (For further details regarding our FDA models and theory, the reader is directed to Ramsay and Silverman, 2005, for example). We note that in the previous discussing regarding the estimation of SV, FDA models were NOT used as their smoothness is believed to mask the true, observed variability in the traffic stream data. However, for other parts of our analysis (see below), the noise in our data is considered as nuisance variability and FDA traces are, in fact, preferable.

Data Traces, Capacity, and Breakdown

Traces derived from FDA models are especially beneficial for extracting capacity data. These traces represent trends of already aggregated data, and the peaks of these smooth curves necessarily correspond to *sustained* maxima which are natural estimates of capacity. Therefore, to help extract realistic signals from the noisy aggregates, daily flow maxima will be extracted from fitted FDA flow traces (smooths). Figure 2 below illustrates the usefulness of the FDA traces as the smooth curve helps to identify a realistic maximal value.

Because the censoring designation of daily flow maxima is dependent on daily breakdowns, and because these values directly relate to fitted capacity distributions, accurate assessment of breakdowns is essential. Like most traffic analyses, we define breakdown to be the transition between freely flowing traffic and congested conditions. This transition, or breakdown, occurs when persistent speeds above a fixed threshold are immediately followed by persistent speeds below the same threshold. This speed threshold was set to 48 mph based on a visually distinction between congested and freely flowing traffic regimes. By applying FDA, we are able to extract interpolated, intermediate speed and flow values based on the smooth representations of the data. Thus, traces address the two issues with our data: they provide precise (in time) estimates while maintaining smoothness. When sustained drops in traffic speeds do occur, associated daily maximal flows are considered observed capacity data. Otherwise, on days where traces do not identify sustained breakdown, associated maxima are censored values. Both these censored and observed capacity values will be used to calibrate the previously defined GEV model.

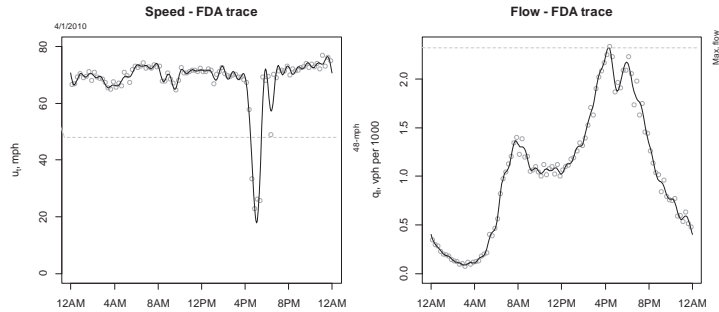


FIGURE 2 Speed (left) and flow (right) FDA traces for April 1, 2010. The small circles represent flow aggregates. Since a sustained speed below 48mph was observed (left), that day’s maximal peak flow (right) was deemed a capacity value.

PROCEDURE (BAYESIAN MODEL FITTING)

Capacity Data

In general, daily maxima (capacity data) have two dominant characteristics: small sample sizes and a high number of censored values. Because of this, and because capacity model-fitting need not be done in real-time, a computational Bayesian approach was employed. Ozguven and Ozbay (2008) concluded that Bayesian estimation is far superior (more efficient) to non-parametric techniques for survival analyses with small samples and substantial amounts of censoring. The OpenBUGS statistically software was used for these applications, and the analysis and manipulation of all OpenBUGS output, the data containing the Bayesian samples of the parameters, was then performed with the R statistical software.

Because there is an absolute limit to the number of vehicles a road may carry, finite upper bounds were assumed for the capacity data, and, consequently, GEV shape parameters were assumed to be negative. The corresponding prior distribution on the shape parameter reflects this limitation. A somewhat diffuse prior distribution for the GEV scale parameter was chosen, but within a realistic range based on previous model-fitting. Similarly, a semi-informative prior distribution was chosen for the location parameter as the center of the distribution can be estimated within a reasonable range. Thus, the scale and shape priors were not over-specified, and we allowed the data to guide these posterior analyses.

Freeway capacities, C , are assumed to be generalized extreme value distributions (GEV)

$$C \sim GEV(\alpha, \beta, \xi)$$

with location, scale, and shape parameters α , β , and ξ , respectively. Collected data are the combined maximum daily traffic flows, each of which is classified as a censored or un-censored, capacity value based on the previously established definition. For the GEV model, the shape (ξ), scale (β), and location (α) parameters are assumed to follow uniform prior distributions on $(-.75, 0)$, $(0, 10)$, and $(3, 10)$, respectively.

Posterior results are based on 5,000 MCMC iterations, the first 2,000 discarded as a ‘burn-in’ period. Convergence and independence from the starting values were checked by CODA (distributions, traces, etc.), the standard tools in such cases. Also, in all cases, starting values for the sampling scheme were generated from the defined prior distributions.

Demand and Stochastic Volatility

Because of the complexity of the stochastic variability model discussed previously, parameter estimation is obtained via Bayesian MCMC methods. Fortunately, the R package ‘stochvol’ (Kastner, 2014) provides an efficient algorithm for such a Bayesian estimation of SV model parameters. To complete the setup, prior distributions for SV parameters μ , ϕ , and λ_n must be established. The parameter μ is assigned a diffuse normal distributions $(N(0, 100))$ as this prior specification is not typically influential on estimation (Kastner, 2014). For ϕ , the prior is specified by beta distribution hyperparameters (5, 1.5). Finally, following Kastner (2014) and Fruerwirth-Schnatter and Wagner (2010), we assign λ_n a centered normal prior with hyperparameter equal to 1. Such a choice of hyperparameter value is not believed to be influential on our final results (Kastner, 2014). For further discussion regarding prior specification and the details of the process, the reader is directed to Kastner and Fruerwirth-Schnatter (2013).

The Bayesian sampling process is repeated for each time period (for each $t = 1, \dots, 96$) to yield distribution estimates for σ_t for each day (for each $i = 1, \dots, 205$). Posterior results are based on 10,000 MCMC iterations, with a ‘burn-in’ period of 1,000 iterations and further thinned. The ‘stochvol’ package converts its output to CODA-compatible objects, so convergence and independence from the starting values were checked by the typical methods (distributions, traces, etc.). In our case, the MCMC iterations indicated convergence to the stationary distribution of the chain.

RESULTS

SV for Individual Days

As stated, the SV model was fitted to each day independently. To observe the effect of flow volatility on the SV parameter estimates, plots of flows and SV estimates were compared for each day. Figures 3 and 4 below presents output for two days of our collection period: both the SV in terms of posterior quantiles of the latent volatilities in percent ($100 * \exp(\frac{\mu_t}{2})$) over time (top) as well as the corresponding daily flow profile on the same time axis.

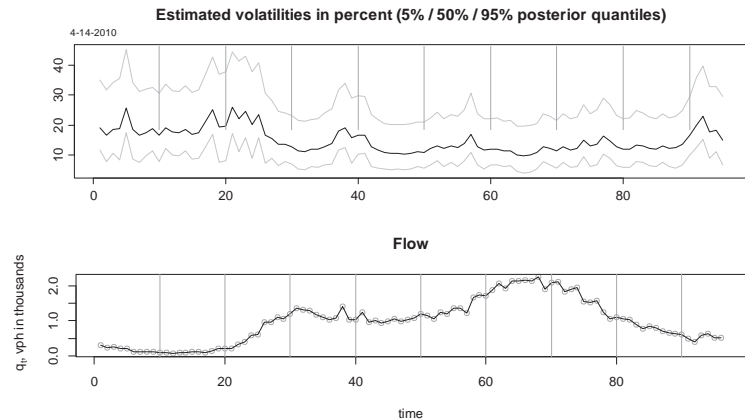


FIGURE 3 Plots of SV in terms of posterior quantiles of the latent volatilities (top) and daily flow values for April 14, 2010.

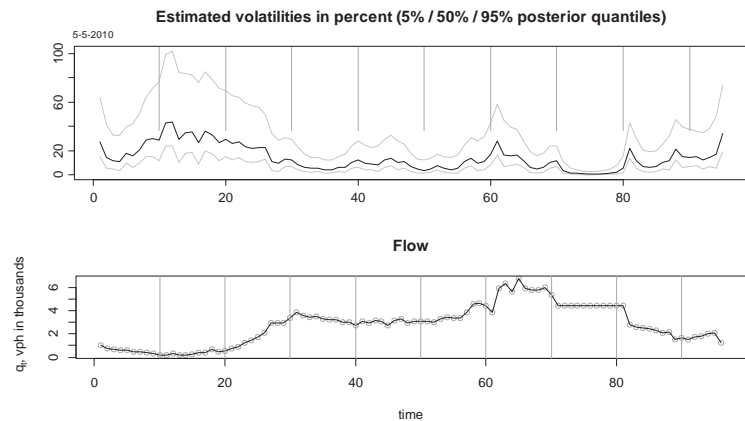


FIGURE 4 Plots of SV in terms of posterior quantiles of the latent volatilities (top) and daily flow values for May 5, 2010.

Figures 3 and 4 are typical of output for most days in that the SV model effectively identifies changes in flow volatility throughout the day. That is, from both plots we can observe strong correspondence between flow volatility and increases (spikes) in SV. Figure 4 is interesting as there was a span of missing flow values on this day (around time step 72-82) which were imputed based on the last observed flow. Observing the SV plot during this same period, we see SV estimates of zero.

SV by Time of Day

As stated previously, the SV was fitted to each day independently. Our procedure then produces an estimate of SV for each of 96 time steps for each of 205 days. Although day-to-day variability is observed in the flow data, most days seem to follow a somewhat similar pattern where volatility increases during morning and evening commuting hours. With this in mind, we have decided to group the daily SV estimates together and identify trends by time of day. More specifically, for the 205 estimate of SV at each of the 96 time steps, we have calculated both the mean confidence bounds. Figure 5 below presents these values as a time series.

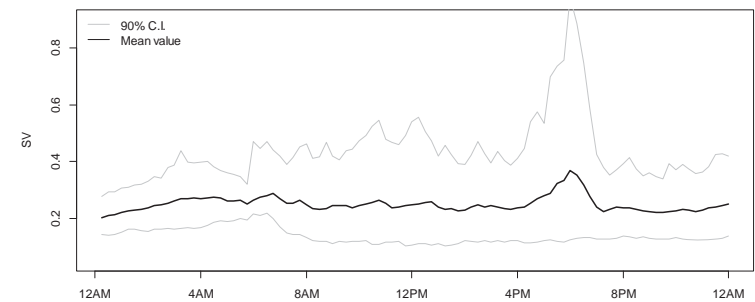


FIGURE 5 Mean stochastic volatility (SV) by time of day (black) along with upper and lower 5% bands (grey).

While the grouping and averaging process across all 205 days does seem to ‘wash away’ much of the variability, we can glean some information from this plot. Clearly, the increase in flow variability during the evening commute (around 6-7PM) is preserved in the SV plot and apparent in the somewhat sharp spike at this time. Since flow volatility is observed during this time on the majority of days, we expect the SV estimates to reflect this behavior. Also, we note that not only are the SV estimates during the evening commuting hours high on average, but the upper confidence band indicates that these values may be very large. Although morning commuting hours also tend to be volatile times, the averaging process seems to mask this behavior. We speculated that morning volatility probably occurs over a wider range of times when consider all days. If this is the case, our procedure would fail to preserve the spikes at

particular time steps and instead give smoother, elevated values. This does appear to be the case as SV estimates tend to be fairly high during the morning hours in comparison to other times of the day.

Breakdown Probabilities

For capacity, for the C model, our Bayesian approach yields posterior vector estimates of model parameters α , β , and ξ . Each vector, then, corresponds to a fitted GEV distribution for capacity. At each quantile of these many distributions, the median value was extracted to produce a single estimate of the GEV distribution form. Similarly, Bayesian models yield estimates for σ_t at all t and for each day. Once these parameters are identified, a simulation-based procedure is used to estimate the probability of breakdown, or breakdown risk. Such an approach takes advantage of the extraordinary wide variety of outcomes observed in the data set.

Before performing our simulation, we first collect flow signal estimates, $q_{i,t}$, for all i and t as identified via FDA traces. Then, we collect all σ_t for each day as identified from our SV model fitting procedure (above). Next we estimate the probability of breakdown by performing the following steps:

- 1) For some $t = t_0$, randomly select a $q_{t_0,i}$ ($i = 1, \dots, 205$).
- 2) For that same $t = t_0$, randomly select a $\sigma_{t_0,j}$ ($j = 1, \dots, 205$) and sample from $N(0, \sigma_{t_0,j})$ to determine d_{t_0} given by $D_{t=t_0} = q_{t=t_0,i} + N(0, \sigma_{t=t_0,j})$.
- 3) Draw a sample from the fitted GEV to obtain c , an estimate for the capacity of the roadway.
- 4) If $d_{t_0} > c$, then record this sample as a breakdown.
- 5) Repeat the process 1,000 times, recording the breakdown state for each sample. The number of breakdowns among the 1,000 samples is denoted $n_{t=t_0}$.
- 6) Calculate the empirical probability for breakdown, $P(D_{t=t_0} \geq C)$, at time $t = t_0$ given by $n_{t=t_0}/1000$. We denote this value, the probability of breakdown at time $t = t_0$, as $p_{t=t_0}$.
- 7) Repeat the process for all t ($t = 1, \dots, 96$) to find all p_t .

As a comparison, to determine the effect of SV on breakdown probability, a second simulation was performed. In this simulation, however, the $q_{t,i}$ were sampled from the raw flow aggregates (as opposed to the FDA signals) and no additional volatility component (SV) was considered. This flow value is then compared to a sampled capacity value (as above) to determine breakdown state. The process is repeated 1,000 times for each time step to identify empirical probabilities of breakdown. Finally, this breakdown probability was compared to that obtained using the SV. Figure 6 below compares the probability estimates for the two sampling procedures by time of day.

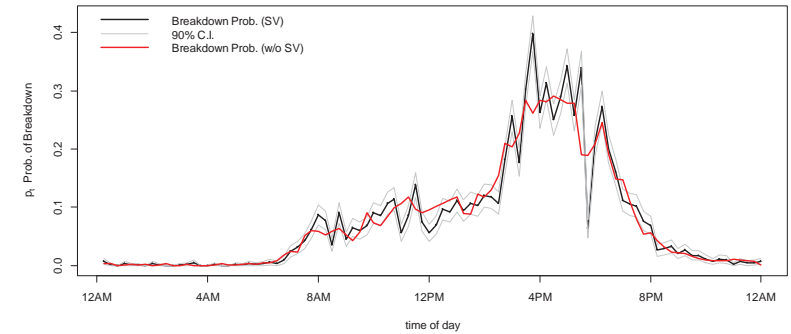


FIGURE 6 Probability of breakdown (p_t) by time of day as calculated by SV model (black) and sampling without considering SV (red). Confidence bounds for SV model predictions are included as well (grey lines).

The effect of introducing noise into our breakdown prediction is quite stark. When modeling the volatility separately and combining it with a flow signal, the prediction of breakdown itself becomes very noisy, much noisier, in fact, than probabilities based on the raw flow data. We observe from Figure 6 that the SV-based predictions of breakdown are not smooth, but tend to fluctuate wildly, especially at known volatile times of the day (morning/evening commutes). This seems to be in keeping with the ‘random’ nature of breakdown occurrence, where breakdowns cannot be predicted based solely on traffic flow.

CONCLUSION AND DISCUSSION

In this paper we have outlined a procedure to predict highway breakdowns based on flow values. The innovative part of our approach is the independent modeling of flow volatility through the use of a stochastic volatility model, a model form commonly used in economic time-series analyses. Functional data analysis (FDA) is used to extract realistic flow signals by time of day. These flow signals are combined with estimates of SV to produce component estimates of demand. In the end, when compared to estimates based on observed data, our estimation procedure yields dissimilar demands and ultimately dissimilar predictions of breakdown. We have simply introduced a procedure, but, going forward, a more thorough simulation-based study should be performed to evaluate the predictive ability of our approach. That is, many of our conclusions are based on visual analysis, and more thorough statistical testing of our model should be pursued. The application of our study is clear: estimation of probabilistic prediction of breakdown can be a valuable tool for decision-makers and administrators. At the very least, results suggest that volatility plays a role in this breakdown, and that the SV-model is a suitable approach to estimate this volatility.

REFERENCES

- Agyemang-Duah, K and F.L. Hall. Some Issues Regarding the Numerical Value of Freeway Capacity, Highway Capacity, and Level of Service. *International Symposium on Highway Capacity*, Karlsruhe, July, 1991, pp. 1-15.
- Brilon, W., J. Geistefeldt, and M. Regler. Reliability of Freeway Traffic Flow: A Stochastic Concept of Capacity. *Proceedings of the 16th International Symposium on Transportation and Traffic Theory*, College Park, Maryland, 2005, pp. 125-144.
- Brilon, W., J. Geistefeldt, and H. Zurlinden. Implementing the Concept of Reliability for Highway Capacity Analysis. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2027, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 1-8.
- Coles, S. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, London, 2001.
- Eleferiadou, L. and P. Lertworawanich. A Methodology for Estimating Capacity at Ramp Weaves Based on Gap Acceptance and Linear Optimization. *Transportation Research, Part B: Methodological*, Vol. 37B, No. 5, 2002, pp. 459-483.
- Eleferiadou, L., R.P. Roess, and W.R. McShane. Probabilistic Nature of Breakdown at Freeway Merge Junctions. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1484, Transportation Research Board of the National Academies, Washington, D.C., 1995, pp. 80-89.
- Fruehwirth-Schnatter, S. and H. Wagner. Stochastic Model Specification Search for Gaussian and Partial Non-Gaussian State Space Models. *Journal of Econometrics*, 154 (1), 2010, pp. 85-100.
- Geistefeldt, J. Consistency of Stochastic Capacity Estimations. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2173, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 89-95.
- Geistefeldt, J. and W. Brilon. A Comparative Assessment of Stochastic Capacity Estimation Methods. *18th International Symposium on Transportation and Traffic Theory*, Hong Kong, 2009.
- Greenwood, I. D., R.C.M. Dunn, R.R. Raine. Estimating the Effects of Traffic Congestion on Fuel Consumption and Vehicle Emissions Based on Acceleration Noise. *Journal of Transportation Engineering*, Vo. 133, No. 2, 2007, pp. 96-104.
- Hall, F.L. and K. Agyemang-Duah. Freeway Capacity Drop and the Definition of Capacity. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1320, Transportation Research Board of the National Academies, Washington, D.C., 1991, pp. 91-98.
- Hall, F.L., V. Hurdle, and J. Banks. Synthesis of Recent Work on the Nature of Speed-Flow and Flow-Occupancy (or Density) Relationships on Freeways. In *Transportation Research Record:*

- Journal of the Transportation Research Board*, No. 1365, Transportation Research Board of the National Academies, Washington, D.C., 1992, pp. 12-17.
- HCM, Highway Capacity Manual. Transportation Research Board, Washington, D.C., 2000.
- Kastner G. Dealing with Stochastic Volatility in Time Series Using the R Package 'stochvol.' Unpublished paper, 2014.
- Kastner, G and S. Fruehwirth-Schnatter. Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models. *Computational Statistics & Data Analysis*, 76, 2013.
- Laflamme, E. *Extreme Value Theory: Applications to Estimation of Stochastic Traffic Capacity and Statistical Downsizing of Precipitation Extremes*. Ph.D. thesis, University of New Hampshire, 2013.
- Lorenz, M. and L. Eleferiadou. Defining Freeway Capacity as Function of Breakdown Probability. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1776, Transportation Research Board of the National Academies, Washington, D.C., 2001, pp. 43-51.
- Minderhoud, M.M., H. Botma, and P.H.L. Bovy. Assessment of Roadway Capacity Estimation Methods. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1572, Transportation Research Board of the National Academies, Washington, D.C., 1997, pp. 59-67.
- Minderhoud, M.M., H. Botma, and P.H.L. Bovy. An Assessment of Roadway Capacity Estimation Methods. *Transportation and Traffic Engineering Section, Report No. VK 2201. 302/LVV 0920-0592*, Delft University of Technology, Delft, 1996.
- Ozguven, E. E. and K. Ozbay. Nonparametric Bayesian Estimation of Freeway Capacity Distribution from Censored Observations. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2061, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 20-29.
- Ramsay, J.O. and B.W. Silverman. *Functional Data Analysis*, 2nd edition. Springer, New York, 2005.
- Taylor, S.J. Financial Returns Modelled by the Product of Two Stochastic Processes—A Study of Daily Sugar Prices 1679-1691. *Time Series Analysis: Theory and Practice*, No. 1, North-Holland, Anderson, O.D. (Ed.), 1982, pp. 203–226.
- Yang, X. and N. Zhang. The Marginal Decrease of Lane Capacity with the Number of Lanes on Highway. *Proceedings of the Eastern Asia Society for Transportation Studies*, Vol. 5, 2005, pp. 739–749.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Ralph J. Cicerone is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the National Academies of Sciences, Engineering, and Medicine to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at www.national-academies.org.

The **Transportation Research Board** is one of seven major programs of the National Academies of Sciences, Engineering, and Medicine. The mission of the Transportation Research Board is to increase the benefits that transportation contributes to society by providing leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation.

Learn more about the Transportation Research Board at www.TRB.org.



TRANSPORTATION RESEARCH BOARD

500 Fifth Street, NW

Washington, DC 20001

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The nation turns to the National Academies
of Sciences, Engineering, and Medicine for
independent, objective advice on issues that
affect people's lives worldwide.

www.national-academies.org