Research based on high-fidelity NGSIM vehicle trajectory datasets: A review

Technical Report · August 2017

DOI: 10.13140/RG.2.2.11429.60643/1

1 author:

Zhengbing He
Beijing University of Technology

Research based on high-fidelity NGSIM vehicle trajectory datasets: A review

Zhengbing He[1]

Abstract

Next Generation Simulation (NGSIM) program published four high-fidelity trajectory datasets more than ten years ago. Recognizing the great influence of the datasets on transportation research, this paper classifies and reviews the research based on the NGSIM trajectory datasets. Due to the wide existence of relevant literature, only the papers published in the leading journal series of Transportation Research are considered. Those papers are then classified into six subjects, and it is found that the data are mainly employed in five ways. To shed light on the future data usage and collection, limitations of the NGSIM datasets are pointed out, and outlooks of future data collection are presented.

Keywords: Next Generation Simulation, vehicle trajectory, traffic data, data usage, traffic flow modeling, transportation research

Please cite the paper in the following way

Zhengbing He, Research based on high-fidelity NGSIM vehicle trajectory datasets: A review, doi:10.13140/RG.2.2.11429.60643, 2017

[1]Dr. **Zhengbing He** is always interested in solving various transportation problems by combining empirical data with the knowledge of traffic flow theory and intelligent transportation systems. His contacts are as follows
Email: he.zb@hotmail.com,
Webpage: http://zhengbing.weebly.com,
Researchgate page: https://www.researchgate.net/profile/Zhengbing_He, and
ORCID: http://orcid.org/0000-0001-5716-3853

August 24, 2017

Contents

1. Introduction

With the arrival of the 21st century, "big data" has become one of the hottest words that we almost hear every day, and various data seem to be everywhere. However, even in such an era of data explosion, the high-fidelity vehicle trajectory data published in Next Generation Simulation (NGSIM) program ten years ago (U.S. Federal Highway Administration, 2006) still play a unique role in transportation research, due to the facts (1) the NGSIM datasets are ones of the few datasets providing high-fidelity vehicle trajectories, although their time-space coverage is limited; (2) by utilizing the NGSIM datasets, a large amount of research was carried out, making people understand traffic unprecedentedly deeply.

Therefore, it is of interest and significance to know what research was conducted based on the NGSIM datasets. Better understanding the past will shed light on the future to making better use of such data and to further collect useful or complementary data. To the end, the review introduces the research based on the high-fidelity NGSIM trajectory datasets, by classifying the research and pointing out the main usage, limitations of the NGSIM datasets, and outlooks of future data collection. Different from most of literature reviews that organized papers along one subject stream, the review summarizes the papers that are all linked by the same data. Moreover, to the best of the authors' knowledge, this is the first time that the research based on high-fidelity vehicle trajectory data or the NGSIM datasets is specially reviewed.

The rest of the review is organized as follows: Section 2 briefly introduces the NGSIM datasets including the US-101, I-80, Peachtree, and Lankershim datasets; Section 3 presents the rule of selecting the papers that are reviewed here; Section 4 classifies and reviews those papers; Sections 5-7 point out the main usage, limitations of the NGSIM datasets, and outlooks of future data collection, respectively; Sections 8 gives a summary to close the review.

2. NGSIM trajectory datasets

The NGSIM datasets were originally collected by using cameras, and then extracted from the resulting videos. The sampling frequency of the NGSIM trajectory is 0.1 sec, and each sample includes the information such as instantaneous speed, acceleration, longitudinal and lateral positions, vehicle length, vehicle type. The descriptions of the four datasets are given as follows.

The US-101 trajectory dataset was collected on a segment in the vicinity of Lankershim Avenue on southbound US-101 freeway in Los Angeles, California. The segment is approximately 640 m in length, and contains 6 lanes (see Figure 1(a)). The time period of data collection is 45 min, i.e., from 7:50 a.m. to 8:35 a.m. on June 15, 2005.

The I-80 trajectory dataset was collected on a segment of I-80 freeway in Emeryville (San Francisco), California. The segment is approximately 500 m in length, and contains 6 lanes, where the median lane is a high occupancy vehicle (HOV) lane (see Figure 1(b)). The data were collected within two periods, i.e., 15 min ranging from 4:00 p.m. to 4:15 p.m. on April 13, 2005, and 30 min ranging from 5:00 p.m. to 5:30 p.m. on April 13, 2005.

The Peachtree trajectory dataset was collected on a segment of Peachtree Street in Atlanta, Georgia. The arterial segment is approximately 640 m in length, with five intersections (four are signalized and one is not) and two or three through lanes in each direction (see Figure 1(c)). The dataset consists of two 15-min time periods, 12:45 p.m. to 1:00 p.m. and 4:00 p.m. to 4:15 p.m. on November 8th, 2006.

The Lankershim trajectory dataset was collected on a segment of Lankershim Boulevard in the Universal City neighborhood of Los Angeles, California. The segment is approximately 488 m in length, and contains three or four lanes and four signalized intersections (see Figure 1(d)). The time period of data collection is 30 min ranging from 8:30 a.m. to 9:00 a.m on June 16, 2005.

3. Selection of literature reviewed

Initially, we searched for a keyword of "NGSIM" in Google Scholar, and obtained over 1700 results[2]. Then, searched in the homepages of mainstream transportation journals, and obtained total 227 results; see Table 1. The large number of results implies that the popularity and importance of the NGSIM datasets, whereas it makes reviewing all of them difficult. Therefore, to filter the results, we only review the literature published in the journal series from Transportation Research Part A (TR-A) to Transportation Research Part F (TR-F). It is known that the Transportation Research series published by Elsevier are leading journals almost covering all aspects of transportation research. It is believed that the literature published in the series is able to well represent the state of the art, although it is admitted that a large number of important publications may be missed[3].

After carefully reading those papers, those only mentioning "NGSIM" (e.g. when introducing other works or pointing out future research direction) instead of actually using the datasets were removed. Finally, total 71 papers were reviewed here[4]; see Table 1 for the numbers of the reviewed papers in each journal. It is noticed that most of the papers were

[2] Google scholar: http://scholar.google.com. It was searched on March 20, 2017

[3] Considering the fact that the scope of IEEE Transactions on Intelligent Transportation Systems is similar to that of TR-C (Lijun and Yafeng, 2017), the 48 papers published in the transaction are not reviewed here, whereas the generality of the following subject classifications may not lose much.

[4] Note that we only briefly introduce the contribution of the literature, since the purpose of this review is to help readers understand what kind of research was conducted based on the NGSIM datasets. For the detailed model or method, interesting readers may refer to the original papers.

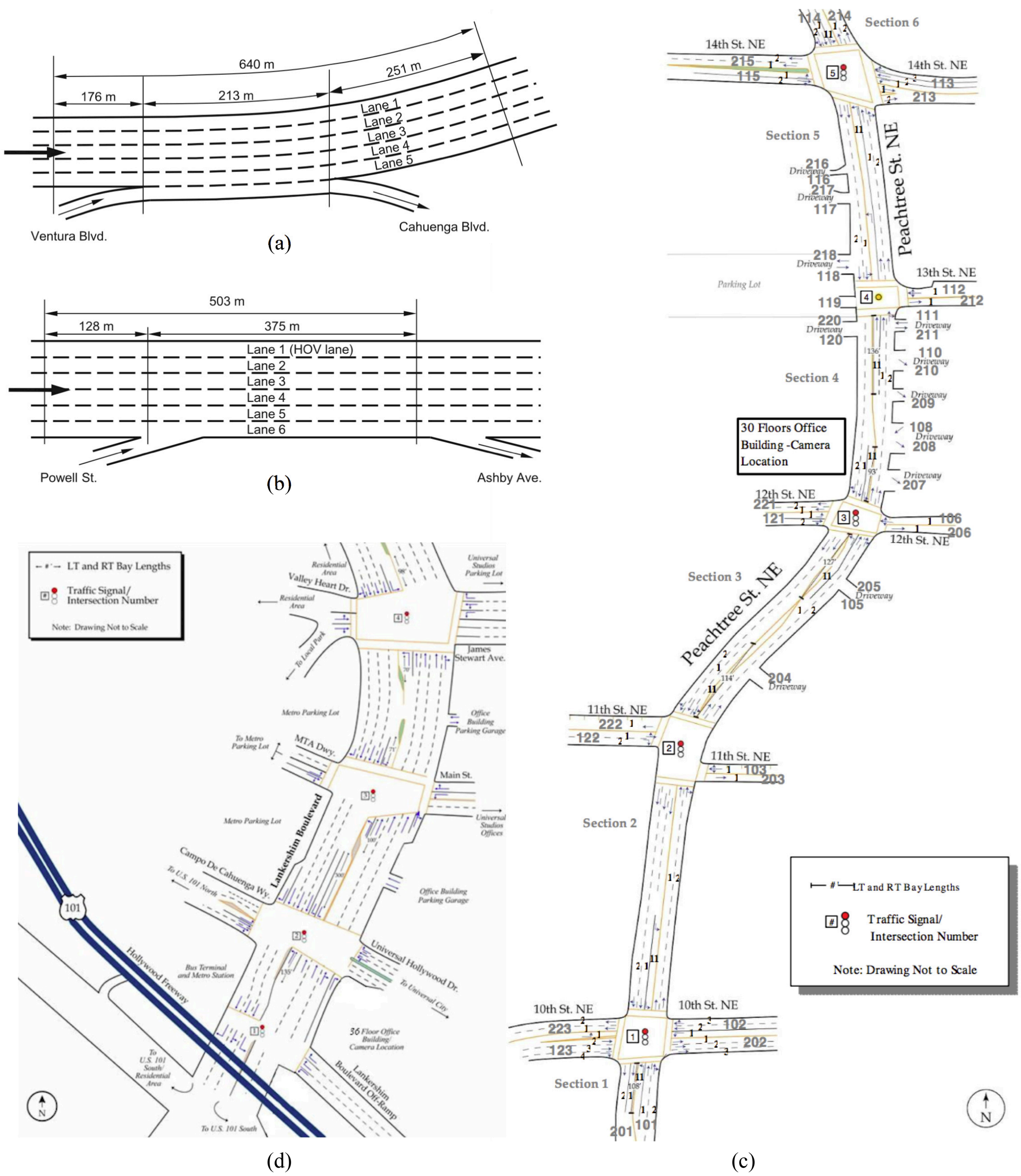

Figure 1: Schematic diagrams of sites collecting the NGSIM trajectories: (a) US-101; (b) I-80; (c) Peachtree; (d) Lankershim.

Table 1: The results of searching for "NGSIM" in transportation journals

Journals	Number of all papers	Number of reviewed papers
Transportation Research Part A: Policy and Practice	0	0
Transportation Research Part B: Methodological	52	40
Transportation Research Part C: Emerging Technologies	56	30
Transportation Research Part D: Transport and Environment	4	0
Transportation Research Part E: Logistics and Transportation Review	0	0
Transportation Research Part F: Traffic Psychology and Behaviour	2	1
IEEE Transactions on Intelligent Transportation Systems	48	—
Computer-Aided Civil and Infrastructure Engineering	7	—
IET Intelligent Transport Systems	8	—
Journal of Intelligent Transportation Systems	10	—
Transportmetrica A: Transport Science	11	—
Transportmetrica B: Transport Dynamics	4	—
ASCE Journal of Transportation Engineering	14	—
Transport Policy	2	—
Networks and Spatial Economics	2	—
Journal of Transport Geography	0	—
Transportation	0	—
Accident Analysis and Prevention	7	—
Total	227	71

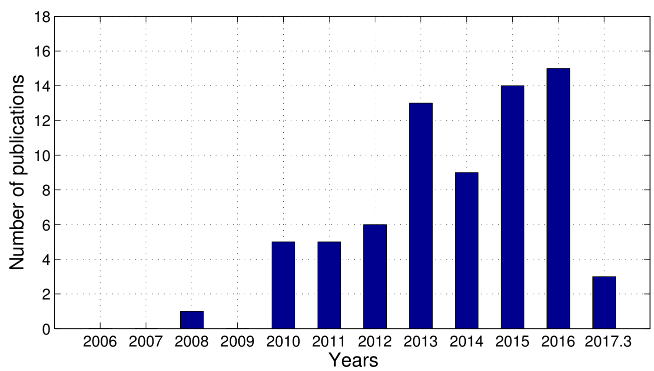

Figure 2: The number of the 71 reviewed papers in each year.

published in TR-B and TR-C, implying that high-fidelity trajectory data were more widely used in (also may be more useful for) the studies of transportation-related methodology and technology instead of transportation-related policy (TR-A), environment (TR-D), logistics (TR-E) and psychology (TR-F). Figure 2 presents the trend of the publications in recent years, indicating that although more than ten years have passed the NGSIM datasets are still popularly used.

4. Subjects of the research based on NGSIM datasets

4.1. Microscopic traffic flow analysis and modeling

Total 25 papers are found under the subject of *microscopic traffic flow analysis and modeling*, which includes five sub-subjects as follows: *model-based car-following modeling*, *data-driven car-following modeling*, *car-following behavior analysis*, *lane-changing modeling*, and *driving strategy development* (refer to Table 2). It is not difficult to understand that *model-based car-following modeling* and *car-following behavior analysis*, which take the advantage of the high resolution of the NGSIM datasets, are two main sub-subjects that include eight and seven papers, respectively. The specific review on each paper and the respective data usage are given as follows.

4.1.1. Model-based car-following modeling

Formulating car-following behaviors in mathematics is called model-based car-following modeling, which is the most prevailing way of modeling driving behaviors. Tordeux *et al.* (2010) proposed an adaptive time gap car-following model, which is able to produce the leader-follower interaction by adjusting the time gap to a targeted safety time, i.e., a function of speed. The US-101 dataset was used to estimate the distribution of the time gaps under different vehicle speeds, i.e., the model parameters for each class of vehicles. Koutsopoulos and Farah (2012) presented a flexible framework to model car-following behavior, which is based on the recognition of driving regimes, such as car-following, free-flow, emergency stopping. In each regime, driver's decisions, such as acceleration, deceleration, and do-nothing, could be made depending on surrounding traffic conditions. Selected leader-follower pairs from the I-80 dataset were used to calibrate both the proposed model and a modified General Motors car-following model (Ahmed, 1999), and those two models were compared using Akaike's information criterion (Akaike, 1974). To reproduce traffic oscillations, Chen *et al.* (2012) proposed a behavioral car-following model, which is based on the empirical finding from the US-101 trajectories that driver behavior before an oscillation is strongly correlated to that during the oscillation. The US-101 trajectories were employed to graphically and statistically analyze car-following behavior, and to calibrate the model parameters. Laval *et al.* (2014) proposed a desired acceleration model and incorporated it into the framework of Newell's simplified car-following model (Newell, 2002). To validate the model, stop-and-go waves were simulated. The simulated pattern of oscillation growth and hysteresis were compared with empirical ones, and a lead vehicle problem (LVP) for 6th follower[5] was tested. After observing the heterogeneity of driving behavior in different spacing-speed states from the US-101 trajectories, Tian *et al.* (2015) assumed oscillating spacing and incorporated it into a cellular automaton to reproduce the empirical observations of Kerner's three-phase theory (Kerner, 2009).

[5]Lead vehicle problem for the nth follower: given leader's complete trajectory and n followers' initial positions and speeds, simulate those followers, and compare the profiles (such as speed and spacing) of the simulated nth follower with those of real nth follower.

Table 2: General information of the papers regarding *microscopic traffic flow analysis and modeling*

Sub-subjects	Papers	Dataset	Journals	Main usage of data
Model-based car-following modeling	Tordeux *et al.* (2010)	US-101	TR-B	Estimating empirical distribution of time gap and model parameters
	Koutsopoulos and Farah (2012)	I-80	TR-B	Calibrating model and comparing with other models
	Chen *et al.* (2012)	US-101	TR-B	Graphically and statistically analyzing car-following behavior and calibrating model
	Laval *et al.* (2014)	US-101	TR-B	Extracting macroscopic traffic patterns as references and testing an LVP for 6th follower
	Tian *et al.* (2015)	US-101	TR-B	Analyzing trajectories for typical driving characteristics
	Wang *et al.* (2011)	I-80	TR-C	Testing an LVP for 1st follower
	Przybyla *et al.* (2015)	US-101, I-80	TR-C	Training model and comparing with simulated trajectories
	Hamdar *et al.* (2015)	I-80	TR-B	Calibrating model using a nonlinear optimization procedure based on a generic algorithm Validating model through a test similar to 2-fold cross-validation
Data-driven car-following modeling	Zheng *et al.* (2013)	US-101	TR-C	Training model (70% data) and validating model (30% data, and an LVP for 8th follower)
	He *et al.* (2015)	US-101, I-80	TR-B	US-101: training model, testing LVPs for 23rd and 30th followers, and extracting macroscopic traffic patterns as references; I-80: showing transferability
	Hao *et al.* (2016)	US-101	TR-C	Three trajectories for model calibration, and four trajectories to test an LVP for 1st follower
Car-following behavior analysis	Chiabaut *et al.* (2010)	I-80	TR-B	Graphically and statistically analyzing leader-follower trajectories
	Chen *et al.* (2012)	US-101	TR-B	Graphically and statistically analyzing leader-follower trajectories
	Chen *et al.* (2014)	US-101	TR-B	Graphically and statistically analyzing trajectories
	Wei and Liu (2013)	US-101	TR-B	5-fold cross-validation and an LVP for 1st follower
	Li *et al.* (2013)	US-101	TR-F	Estimating headway distributions and model parameters
	Taylor *et al.* (2015)	I-80	TR-B	Estimating the distributions of model parameters
	Hamdar *et al.* (2016)	I-80	TR-C	Building simulation environments with empirical traffic
Lane-changing modeling	Laval and Leclercq (2008)	I-80	TR-B	Illustrating traffic phenomena
	Zheng *et al.* (2013)	US-101, I-80	TR-B	Illustrating lane-changing process, calibrating and validating models
	Talebpour *et al.* (2015)	US-101	TR-C	Calibrating and validating lane-changing model
	Balal *et al.* (2016)	US-101, I-80	TR-C	I-80: calibrating and validating model; US-101: testing for transferability
Driving strategy development	Yang and Jin (2014)	US-101	TR-C	Building simulation environments with an empirical leader and an initial condition
	Gong *et al.* (2016)	I-80	TR-C	Building a simulation environment with an empirical leader
	Tak *et al.* (2016)	US-101	TR-C	Providing macroscopic and microscopic data, calculating collision risk, and analyzing system performance

Three safety-related car-following studies are founded and introduced as follows. To understand vehicle-to-vehicle dynamic interactions and prevent rear-end collisions, Wang *et al.* (2011) proposed a driver's safety approaching behavioral model by considering the variability of a follower's speed and spacing to its leader. To validate the model, an empirical leader was selected from the I-80 dataset, and the LVP was tested for 1st follower. To estimate risk effects of driving distraction, Przybyla *et al.* (2015) proposed a dynamic errorable car-following model by using a dynamic time warping algorithm. Leader-follower car-following pairs were extracted from both the US-101 and I-80 trajectory datasets and used to train the dynamic time warping algorithm. Subsequently, the simulated trajectories were compared with real ones to analyze the risk effects. Hamdar *et al.* (2015) proposed a stochastic car-following model with some important behavioral and psychological considerations, such as subjective utilities and dis-utilities for acceleration and deceleration, risk taking. To calibrate the stochastic model, a nonlinear optimization procedure based on a generic algorithm (Hamdar *et al.*, 2009) was used. A test analog to 2-fold cross-validation[6] was conducted by splitting the I-80 dataset into two folders, i.e., one is between 4:00 p.m. to 4:15 p.m., and the other is between 5:00 p.m. to 5:15 p.m.

4.1.2. Date-driven car-following modeling

Compared with the model-based modeling, date-driven car-following modeling is mainly based on artificial intelligence without mathematics directly describing car-following behaviors. Zheng *et al.* (2013) developed a neural-network-based car-following model by building and incorporating a neural network for instantaneous reaction delay. 70% of the US-101 data were used to train the back-propagation algorithm, and the remaining 30% were used to test an LVP for 8th follower. He *et al.* (2015) proposed a K-Nearest-Neighbor based car-following model, whose four inputs are leader's moving distances and follower's space headways in the latest two time steps, and whose output is follower's moving distance. The US-101 dataset was used to train the model, and the I-80 dataset was compared to show the transferability of the proposed model. To validate the model, LVPs for 23rd and 30th followers were tested, and important traffic characteristics such as wave speed, fundamental diagrams were extracted from the US-101 dataset as references to compare with the simulated ones. Hao *et al.* (2016) proposed a fuzzy logic-based car-following model with a five-layer structure, i.e., Perception-Anticipation-Inference-Strategy-Action. Seven trajectories were selected from the US-101 trajectories on the leftmost lane, three of which were used

[6] k-fold cross-validation: the original sample is randomly partitioned into k equal sized folders. Of the k folders, a single folder samples is retained as the validation data for testing the model, and the remaining $m-1$ folder samples are used as training data. The cross-validation process is repeated k times (the folds), with each of the k folders used exactly once as the validation data. The k results from the folds can then be averaged to produce a single estimation.

for model calibration, and four of which were used for validation by testing an LVP for 1st follower.

4.1.3. Car-following behavior analysis

Comparing with car-following modeling, the papers in this sub-subject more focus on analyzing car-following behaviors and finding microscopic driving characteristics. Chiabaut *et al.* (2010) carefully analyzed the parameters of Newell's car-following model (Newell, 2002) in congested traffic conditions, and then established relations between stochastic Newell's model with heterogeneous drivers and its associated macroscopic pattern. Chen *et al.* (2012) studied traffic hysteresis from a behavioral perspective, and found that the occurrence and type of traffic hysteresis is closely correlated with driver behavior when experiencing traffic oscillations. Moreover, Chen *et al.* (2014) further investigated on the formation of traffic phenomena from the car-following perspective, including how periodic oscillation forms, how driver characteristics contribute to capacity drop, etc.

All the literature carried on their studies mainly by analyzing the leader-follower trajectories graphically and statistically, in particular on the basis of Newell's car-following theory (Newell, 2002), such as estimating the distribution of wave speed, plotting time-space trajectory and flow-density diagrams.

Wei and Liu (2013) employed a self-learning support vector regression approach to investigate the asymmetric characteristic in car-following as well as its impacts on traffic flow evolution. 5-fold cross-validation was conducted to train and validate the approach, and an LVP for 1st follower was tested for model validation. Li *et al.* (2013) proposed an asymmetric stochastic extension of the Tau Theory (Lee, 1976) to explain the phenomenon that vehicle's headway follows a certain log-normal type distribution within different speed ranges, implying that the physiological Tau characteristics implicitly affect human driving behavior and the resulting traffic dynamics. The US-101 dataset was used to estimate the distributions of headways as well as the parameters of the Tau Theory. To examine intradriver heterogeneity, Taylor *et al.* (2015) proposed a dynamic time warping algorithm to analyze vehicle trajectories and car-following behavior. The proposed dynamic time warping algorithm was employed to extract the distributions of the parameters in a stochastic extension of Newell's car-following model (Newell, 2002) from the I-80 dataset. To quantify driver behavior under different roadway geometries and weather conditions, Hamdar *et al.* (2016) extended the acceleration modeling framework based on the Prospect Theory (Kaneman and Tversky, 1979), and conducted driving experiments using a driving simulator. Typical I-80 trajectories were selected to build the simulated traffic environment (at least five vehicles were in front of and behind the lead vehicle, respectively), in order to make it as generic as possible.

4.1.4. Lane-changing modeling

Lane-changing is more complicated driving behavior including simultaneous longitudinal and lateral movements. Four papers are founded to model the lane-changing behavior in a microscopic perspective. Laval and Leclercq (2008) introduced a framework to capture the relaxation phenomena commonly observed near congested on-ramps, i.e., vehicles take short spacings when entering a freeway, but "relax" to more comfortable spacings shortly thereafter. The I-80 data were plotted in time-space trajectory and flow-density diagrams to illustrate the relaxation phenomena. Zheng *et al.* (2013) investigated the anticipation and relaxation processes in lane-changing, and offered an extension to Newell's car-following model (Newell, 2002) to describe a regressive effect. The trajectories of lane-changers were selected and graphically analyzed, and the parameters (representing anticipation and relaxation) in the extension of Newell's car-following model were calibrated and statistically investigated. To validate the model, the selected lane-changing data were split into two subsets, and their modeling results were compared. Talebpour *et al.* (2015) proposed a lane-changing model in a connected-vehicle environment, based on a game theory (two-person non-zero-sum non-cooperative game) approach that endogenously accounts for the information flow. Mandatory and discretionary lane-changing behaviors were identified in the US-101 dataset. 15 min (7:50 a.m. to 8:05 a.m.) data were used to calibrate the model, and another 15 min (8:05 a.m. to 8:20 a.m.) data were used to validate it by calculating estimation errors. Based on a fuzzy inference system, Balal *et al.* (2016) presented a binary decision model to determine if it is time to execute a lane change. Randomly-selected 70% trajectories of the I-80 dataset were used to calibrate the fuzzy inference system, and the remaining 30% were used to test the results. In addition, the US-101 dataset was served to test the calibrated model in order to show its transferability.

4.1.5. Driving strategy development

To reduce driving emissions or to increase driving safety, driving strategies or driver assistance systems utilizing new technologies are proposed in the following three papers. Yang and Jin (2014) developed a distributed cooperative green driving strategy based on inter-vehicle communications, in order to smooth traffic flow and lower pollutant emissions and fuel consumption in stop-and-go traffic. Ten trajectories with a high standard deviation in speed were selected and combined as a leading trajectory, and the speed profiles of the follower before and after applying the proposed strategy were compared to demonstrate the effect of the proposed strategy. In addition, the initial locations of the vehicles in the US-101 dataset were taken as an initial condition to build a simulation scenario to test the effect of the strategy with different penetration rates and communication delays. Gong *et al.* (2016) proposed a car-following control scheme for a platoon of connected and autonomous vehicles, which was modeled as an interconnected dynamic system subject to acceleration, speed, and safety distance constraints. An oscillating vehicle trajectory was selected from the I-80 dataset and taken as an empirical leader of a simulation scenario. Tak *et al.* (2016) developed a hybrid collision warning system that was able to utilize the integrated information of macroscopic loop detector data and microscopic smartphone data. The US-101 trajectories were taken as the microscopic data, and virtual loop detectors were set to calculate macroscopic data. The collision risk resulting from three collision warning systems was calculated and compared, and two detailed vehicle trajectories were selected to analyze the performance of the systems.

4.2. Mesoscopic and macroscopic traffic flow modeling

Only five papers are found under the subject of *mesoscopic and macroscopic traffic flow modeling* (refer to Table 3). Chiu *et al.* (2010) proposed a vehicle-based mesoscopic traffic simulation model that explicitly considers the anisotropic property of traffic flow (i.e., vehicles mostly react to other vehicles that are in front of them (Daganzo, 1995)) into the vehicle state update at each simulation step. The I-80 dataset was employed to calibrate the proposed model by minimizing the introduced variable of a speed influencing region. Piccoli *et al.* (2015) presented several data fusion schemes to incorporate vehicle trajectory data into a second-order phase transition model, and evaluated the estimation accuracy of first-order variables (such as speed and density) and second-order variables (such as acceleration and emission), by using mobile sensor data with various penetration rates and sampling frequency. A standard k-fold cross-validation was conducted, and the value of k was set based on the penetration rates. Qian *et al.* (2017) developed a macroscopic heterogeneous traffic flow model by considering interplay of multiple vehicle classes, and introduced an intuitive computational procedure to capture mixed vehicular flow propagation and shock wave formation. To validate the model, an initial value problem[7] (IVP) based on the I-80 dataset was solved.

Two papers are related to the macroscopic traffic models taking lane-changing into account. Jin (2010) proposed a macroscopic kinematic wave model to capture bottleneck effects and to aggregate traffic dynamics of lane-changing traffic. Jin (2013) developed a multi-commodity (i.e., weaving and non-weaving vehicles) behavior Lighthill-Whitham-Richards model of lane-changing traffic flow. A lane-changing fundamental diagram was introduced, which is determined by both car-following and lane-changing characteristics as well as road geometry and traffic composition. Both the studies calibrated the newly introduced concepts, i.e., the lane-changing intensity and the lane-changing fundamental diagram, by measuring

[7]Initial value problem: given empirical boundary and initial traffic conditions extracted from empirical data, estimate intermediate traffic states using the proposed model and then compare them with the real traffic states.

empirical traffic variables from the I-80 dataset, such as on-ramp flow rate, lane-changing times.

Table 3: General information of the papers regarding *mesoscopic and macroscopic traffic flow modeling*

Papers	Dataset	Journals	Main usage of data
Chiu *et al.* (2010)	I-80	TR-B	Calibrating model by formulating an optimization problem
Piccoli *et al.* (2015)	I-80	TR-C	k-fold cross-validation
Qian *et al.* (2017)	I-80	TR-B	Validating model: inputting as boundary and initial conditions of IVP
Jin (2010)	I-80	TR-B	Calibrating model, i.e., the lane-changing intensity
Jin (2013)	I-80	TR-B	Calibrating model, i.e., the lane-changing fundamental diagram

4.3. Traffic-related estimation and prediction

The subject of *traffic-related estimation and prediction* is another main subject in utilizing the NGSIM datasets. Total 26 papers are found under the subject, and five sub-subjects are further classified as presented in Table 4 and introduced as follows.

4.3.1. Macroscopic traffic variables estimation

In this sub-subject, macroscopic traffic variables, such as traffic flow, speed, are estimated based on the NGSIM datasets, and the fundamental diagrams are constructed. Coifman (2015) presented a method to estimate the fundamental diagrams with no need of seek out stationary conditions. Vehicle length was found to be the key. As a complement of loop detector data, the I-80 dataset with higher resolution was analyzed by using the proposed method to increase the credibility of the method. Siqueira *et al.* (2016) proposed an alternative stochastic model for the fundamental diagrams, by introducing a stochastic transport model with discrete speed spectrum. The I-80 dataset was used to calculate model parameters and to estimate the empirical fundamental diagrams, which were then compared with the estimation results made by the proposed model. Based on Newell's car-following model (Newell, 2002), Jabari *et al.* (2014) proposed a stochastic version of the macroscopic traffic flow speed-density relation, which allows to investigate the impact of driver heterogeneity on macroscopic traffic flow relations. The first 15-min (4:00 p.m. to 4:15 p.m.) I-80 data were used to estimate the distributions of model parameters, and the other 30-min (5:00 p.m. to 5:30 p.m.) data were used to plot an empirical speed distribution, which was then compared with the simulation results. Wu and Coifman (2014) proposed a length-based vehicle classification method from dual-loop detectors by considering vehicle acceleration in congested traffic. Through setting virtual loop detectors, the I-80 dataset was used to evaluate the performance of the proposed vehicle classification method. The vehicle length information contained in the high-fidelity NGSIM datasets makes evaluating the vehicle length-based study possible.

Table 4: General information of the papers regarding *macroscopic traffic variables estimation*

Sub-subjects	Papers	Dataset	Journals	Main usage of data
Macroscopic traffic variables estimation	Coifman (2015)	I-80	TR-B	Being analyzed by using the proposed method as a complement of loop detector data
	Siqueira *et al.* (2016)	I-80	TR-B	Calibrating model parameters and estimating the referred fundamental diagram
	Jabari *et al.* (2014)	I-80	TR-B	Calibrating the distributions of model parameters using the (first 15 min) I-80 data, and validating model by comparing with speed distributions (other 30 min)
	Wu and Coifman (2014)	I-80	TR-C	Evaluating performance of the proposed method after setting virtual loop detectors
Macroscopic traffic phenomena estimation and analysis	Laval (2011)	US-101	TR-B	Applying the respective methods to extract and measure macroscopic traffic phenomena or the phenomena-related variables
	Ahn *et al.* (2013)	US-101, I-80	TR-C	
	Zheng *et al.* (2011a)	US-101	TR-B	
	Zheng *et al.* (2011b)	US-101, I-80	TR-B	
	Blandin *et al.* (2013)	I-80	TR-B	
	Oh and Yeo (2015)	US-101	TR-B	
	Li *et al.* (2014)	US-101	TR-B	
Traffic states estimation	Herrera and Bayen (2010)	US-101	TR-B	Calculating ground-truth traffic states and simulating various probe vehicle data
	Deng *et al.* (2013)	I-80	TR-B	Calculating ground-truth traffic states
	Bucknell and Herrera (2014)	US-101	TR-C	Calculating ground-truth traffic states and simulating various probe vehicle data
	Argote-Cabañero *et al.* (2015)	Peachtree	TR-C	Calculating ground-truth measures of effectiveness and simulating connected vehicle data
Travel time estimation	Ramezani and Geroliminis (2012)	Peachtree	TR-B	Calculating ground-truth travel time for evaluating estimation results
	Feng *et al.* (2014)	Peachtree	TR-C	
Intersection traffic-related estimation	Qi *et al.* (2013)	Lankershim	TR-C	Calibrating the triangular fundamental diagram and the proposed model
	Srivastava *et al.* (2015)	Lankershim	TR-B	Calibrating the triangular fundamental diagram
	Sun and Ban (2013)	Peachtree	TR-C	Validating model, since the data reflect complete ground-truth traffic
	Hao *et al.* (2013)	Peachtree	TR-C	
	Hao *et al.* (2014)	Peachtree	TR-B	
	Sun *et al.* (2013)	Peachtree	TR-B	
	Yang *et al.* (2016)	Lankershim	TR-C	
	Lee *et al.* (2015)	Lankershim	TR-C	Calibrating model: first 10-cycle data; validating model: remaining 10-cycle data
	Lee and Wong (2017)	Lankershim	TR-B	Validating model

4.3.2. Macroscopic traffic phenomena estimation and analysis

This sub-subject studies on macroscopic traffic phenomena, such as hysteresis, oscillations, capacity drop, based on the NGSIM datasets. Laval (2011) found a new shape for the well-known hysteresis phenomenon in traffic flow after aggregating time-space trajectories along the wave direction by using Edie's definitions (Edie, 1963). Ahn *et al.* (2013) later investigated the hysteresis as vehicles experienced stop-and-go waves, and it was found that the hysteresis takes place less frequently and in smaller amplitude than previously thought. Zheng *et al.* (2011a) applied wavelet transform to analyze important features related to bottleneck activations and traffic oscillations in congested traffic in a systematic manner. Zheng *et al.* (2011b) subsequently demonstrated a way of using wavelet transform to identify the formation and propagation of the stop-and-go waves. Blandin *et al.* (2013) presented a phase transition model of non-stationary traffic to model complex macroscopic traffic phenomena such as hysteresis and phantom jams. Oh and Yeo (2015) analyzed the capacity drop from a microscopic perspective, and found several factors that may trigger the drop, such as driver's tendency to take a large headway after passing stop-and-go waves. Using their previously proposed describing-function based approach (Li *et al.*, 2012), Li *et al.* (2014) estimated fuel consumption and emission from traffic oscillations and explored vehicle control strategies to smooth traffic.

All the above studies regarding *macroscopic traffic phenomena estimation and analysis* mainly applied the respective methods to extract and measure macroscopic traffic phenomena or the phenomena-related variables, such as traffic flow, capacity, and headway, from the NGSIM trajectories.

4.3.3. Traffic states estimation

This sub-subject focuses on estimating traffic states, in particularly estimating intermediate traffic states given downstream and upstream boundaries. Given downstream and upstream traffic states to estimate intermediate states, Herrera and Bayen (2010) presented a method to incorporate mobile probe sensor data into a freeway traffic flow model. This method can work even when data are not available for the on- and off-ramps. Deng *et al.* (2013) extended Newell's three-detector problem (Newell, 1993), and presented a stochastic traffic state estimation method using multiple data sources, such as loop detector and floating car data. Bucknell and Herrera (2014) studied on the accuracy of traffic states estimation (especially in reconstructing traffic speed) using various penetration rates and sampling frequency of mobile sensors. Argote-Cabañero *et al.* (2015) studied on the estimation of measures of effectiveness (such as average speed, number of stops, delay) for traffic operations under a connected-vehicle environment, and determined the minimum connected-vehicle penetration rate to estimate the measures of effectiveness. Since the Peachtree dataset is lack of saturated conditions, only undersaturated conditions were studied using the empirical data in Argote-Cabañero *et al.* (2015).

Taking the advantage that the NGSIM datasets provide a complete picture of the monitored traffic, the papers under this sub-subject employed the NGSIM datasets to calculate ground-truth traffic states or variables for model or method evaluation, or to simulate probe vehicle data or connected vehicle data with different penetration rates and/or sampling frequency.

4.3.4. Travel time estimation

Two papers studying on travel time estimation are found. Given probe vehicles' travel times, Ramezani and Geroliminis (2012) estimated arterial travel time by applying a Markov chain procedure to integrate travel time correlation of routes successive links. To estimate arterial travel time using probe vehicle data, Feng *et al.* (2014) employed mixtures of normal distributions to approximate link travel time distributions, and then estimated real-time travel time using probe vehicle travel time based on the Bayes Theory.

Similar to the studies of estimating traffic states, the data usage here is also to provide ground-truth travel time for evaluating estimation results.

4.3.5. Intersection traffic-related estimation

All the papers focusing on estimating traffic states or variables at intersection, such as discharge flow, queue length at intersection, are placed in this sub-subject. Qi *et al.* (2013) established the relationship between discharge flow and (efficiency- and objective-driven) lane-changing behavior at signalized intersections, and proposed an enhanced Cell Transmission Model by incorporating the lane-changing behavior. To model the realistic discharge flow rate and headway features at signalized intersections, Srivastava *et al.* (2015) presented a modified Cell Transmission Model by substituting the traditional demand function with a linearly decreasing function, and solved it under various Riemann problem scenarios.

Since both of the above studies were based on the Cell Transmission Model and focused on the traffic at signalized intersections, the Lankershim dataset was thus employed to calibrate the triangular fundamental diagram simply by a regression method or observation. In Qi *et al.* (2013), a driving behavioral parameter in the proposed model was also calibrated.

Sun and Ban (2013) presented optimization- and delay-based models based on a variational formulation of traffic flow (Daganzo, 2005) to reconstruct vehicle trajectories from mobile traffic sensors at arterial intersections. Hao *et al.* (2013) proposed a three-layer Bayesian Network model to describe the stochastic intersection flow, by capturing the relationship between the arrival and departure processes and vehicle indices, which is a newly introduced concept in a cycle at a signalized intersection. Hao *et al.* (2014) presented a Bayesian Network based model to estimate the cycle-by-cycle queue length distribution of a signalized intersection, by using sample travel times collected from mobile sensors. To balance data needs for transportation modeling and privacy protection, Sun *et al.* (2013) developed a virtual trip lines zone-based system that is able to ensure an acceptable level of privacy and result in satisfactory results of transportation applications. Yang *et al.* (2016) improved a signal control algorithm developed for connected vehicles (Ilgin Guler *et al.*, 2014) in several aspects, such as integrating three different stages of technology development, developing a heuristic method to switch the signal controls, incorporating trajectory design for automated vehicles.

All the above studies are based on partial traffic data (such as collected by mobile sensors or in virtual trip lines zones), and thus the advantage of the NGSIM trajectory data, i.e., providing complete traffic information, is taken to validate the proposed estimation methods, which is similar to the data usage in the estimations of traffic states and travel time. It is worth mentioning that to overcome the insufficiency issue of the NGSIM datasets, Hao *et al.* (2014) ran 20 replicas of the estimation method for each cycle under each penetration rate.

To estimate lane-based queue lengths at intersection in real time, Lee *et al.* (2015) developed discriminant models to identify critical issues, such as where there is a residual queue and estimating downstream arrivals for each lane. Lee and Wong (2017) proposed a group-based approach to estimate lane-based incremental queue accumulations, and presented a control delay model to predict temporal and spatial factors in future incremental queue accumulations as well as to produce the most appropriate time windows for a rolling horizon procedure.

Both the above papers conducted their lane-based studies by utilizing the high resolution of the NGSIM datasets collected at intersection, i.e., extracting lane-based queue lengths. In addition, Lee *et al.* (2015) converted the Lankershim trajectories to loop detector data by setting virtual downstream and upstream loop detectors. Then, the data collected during the first 10 cycles were used to calculate model parameters, and the remaining 10-cycle data were used to validate the model by comparing empirical and simulated queue lengths.

4.4. Traffic flow model calibration

Making use of the NGSIM datasets, total eight papers study on the *traffic flow model calibration* as shown in Table 5.

Table 5: General information of the papers regarding *traffic flow model calibration*

Papers	Dataset	Journals	Main usage of data
Li *et al.* (2012)	US-101	TR-B	
Rhoades *et al.* (2016)	US-101	TR-B	
Kim *et al.* (2013)	US-101	TR-C	
Vieira da Rocha *et al.* (2015)	I-80	TR-D	Being an empirical sample to
Li *et al.* (2016)	US-101	TR-C	demonstrate the proposed method
Durrani *et al.* (2016)	US-101	TR-C	
Zhong *et al.* (2016)	I-80	TR-C	
Sopasakis and Katsoulakis (2016)	US-101	TR-B	

By using vehicle trajectory data with extracted frequency-domain characteristics, Li *et al.* (2012) proposed a systematic framework to validate the describing-function approach (Li *et al.*, 2012) that is an analytical approach able to predict traffic oscillation propagation for a general class of car-following models. To calibrate nonlinear car-following laws based on leader-follower trajectories, Rhoades *et al.* (2016) proposed a calibration method that takes into account not only driver's car-following behavior but also time- and frequency-domain properties of vehicle trajectories. Kim *et al.* (2013) proposed a robust algorithm called the expectation-maximization to calibrate a General Motors car-following model (Chandler *et al.*, 1958) with random coefficients reflecting drivers' heterogeneity. Vieira da Rocha *et al.* (2015) studied on the effectiveness of goodness-of-fit indicator-based calibration method of car-following models in estimating environment-related factors, such as fuel consumption, nitrogen oxide and particulate matter emissions. To better calibrate car-following models, Li *et al.* (2016) proposed a global optimization algorithm that integrates global direct search and local gradient search to find the optimal solution in an efficient manner. Durrani *et al.* (2016) calibrated the driving behavior parameters for cars and heavy vehicles in the Wiedemann 99 vehicle-following model (Aghabayk *et al.*, 2013), and demonstrated the significant effect of the leader class on the flower's behavior in car-following model calibration. Zhong *et al.* (2016) proposed a cross-entropy calibration method to identify parameters of deterministic car-following models, by formulating it as a stochastic optimization problem. Moreover, a probabilistic sensitivity analysis algorithm was introduced to identify the most important parameters to simply the calibration process. Sopasakis and Katsoulakis (2016) proposed a dynamic model parameterization approach to appraise traffic flow models and to optimize their performance against time-series traffic data and prevailing conditions by analyzing perturbations of vehicle trajectories. A mathematical method that quantifies traffic information loss was additionally presented.

All the studies regarding *traffic flow model calibration* mainly took the NGSIM dataset as an sample to apply and demonstrate their proposed methods.

4.5. Vehicle trajectory data cleaning

As early pointed out in Punzo *et al.* (2011), there are errors in the NGSIM datasets, in particular in the instantaneous data of speed and acceleration. Therefore, five papers aiming at (trajectory) data cleaning are found as presented in Table 6 and as introduced as follows.

To inspect trajectory data accuracy and reduce noise, Punzo *et al.* (2011) designed quantitative methods including jerk analysis, consistency analysis, and spectral analysis, and applied them on the complete four NGSIM datasets. Cleaned NGSIM datasets were subsequently published and expected to be a benchmark for trajectory data quality. Zheng and Washington (2012) studied on the issue of selecting an optimal wavelet to detect irregular structures and transient phenomena in traffic data, and recommended the Mexican hat

Table 6: General information of the papers regarding *vehicle trajectory data cleaning*

Papers	Dataset	Journals	Main usage of data
Punzo *et al.* (2011)	US-101, I-80 Peachtree, Lankershim	TR-C	Being an empirical sample to demonstrate the proposed method
Zheng and Washington (2012)	US-101	TR-C	
Montanino and Punzo (2015)	I-80	TR-B	
Fard *et al.* (2017)	I-80	TR-C	
Zheng and Su (2016)	US-101	TR-B	

wavelet for cleaning traffic and vehicular data. Montanino and Punzo (2015) proposed a traffic-informed method to denoise and reconstruct vehicle trajectories. A simulation-based framework was proposed to verify that the reconstructed trajectories were closer to the real ones (that was actually unknown) than the collected ones. For the similar purpose, Fard *et al.* (2017) proposed a two-step technique based on wavelet analysis, i.e., first identifying and modifying outliers, and then eliminating them by applying a wavelet-based filter. To fill the gap that the types of noise in traffic data were ignored, Zheng and Su (2016) proposed a compressed sensing theory based algorithm to recovery traffic data with Gaussian measurement noise, partial data missing, and corrupted noise. Moreover, Markov random field and total variation regularization were used to improve the accuracy of traffic state estimation.

Since there are various errors in the NGSIM trajectory data, the NGSIM trajectories were naturally taken as study objects in the studies regarding *vehicle trajectory data cleaning*.

4.6. Vehicular Ad Hoc Network-related studies

Although Vehicular Ad Hoc Network (VANET) has not widely deployed in practice, two papers are found to be related to VANET (see Table 7).

Table 7: General information of the papers regarding *VANET-related studies*

Papers	Dataset	Journals	Main usage of data
Baiocchi (2016)	US-101, I-80, Peachtree	TR-B	Being as VANET testbeds
Du *et al.* (2016)	US-101	TR-C	

Aiming at the applications based on VANET, Baiocchi (2016) presented an analytical model of message coverage distance and delivery delay with timer-based dissemination protocols, which is able to evaluate the trade-off between delay due to timers and covered distance. To capture information spreading dynamics via VANET, Du *et al.* (2016) developed an information-traffic coupled cell transmission model, which discretizes a road segment into a number of cells, and mathematically captures the inner-cell and inter-cell movements of information front and tail.

In the studies, the NGSIM datasets were treated as VANET testbeds to evaluate model performance, by assuming the VANET applications have been practically deployed.

5. Usage

In general, the NGSIM datasets have two main advantages for transportation research. The first one is its *high resolution*, which allows researchers to investigate very detailed driving behaviors and to calibrate or estimate microscopic behavioral parameters and variables. The second one is its *completeness* in reflecting traffic conditions, which provides researchers with 100 percent and ground-truth traffic during the collection period and at the collection locations. Taking the advantages, those studies mainly utilize the NGSIM trajectory data in the following ways (also refer to Table 2-7):

• Calibrating or training traffic flow models. To calibrate observable traffic parameters, such as the fundamental diagrams and a distribution of headways, we can directly estimate them from the time-space trajectories. For unobservable parameters, the calibration is usually to solve an optimization problem, in which the decision variables are the parameters of a model (such as a car-following model) and the objective function characterizes the difference between empirical vehicle movements and their simulated correspondences. In addition, for the data-driven models without parameters, the NGSIM datasets are used to train the model. Note that to eliminate the correlation with model validation, it is usual that only a part of the dataset is employed to calibrate models, and the other part is used to validate models.

• Validating models. Three prevailing methods of validating models are as follows. (1) An LVP test for nth follower, in particular for validating car-following models. (2) (k-fold) cross-validation. The method is usually used for the model that needs not only validation but also calibration. (3) Direct comparison with ground truth (such as by calculating error magnitudes), including comparison with ground-truth traffic states, ground-truth traffic phenomena, etc. The validation method is widely applied in the studies related to mobile sensors, and the studies of estimating traffic state and travel time. For mobile sensor-related studies, the NGSIM datasets not only provide ground-truth traffic, but its completeness also allows researchers to carry on experiments with different penetration rates of mobile sensors. Likewise, for traffic state and travel time estimations, the completeness allows researchers to flexibly set or select any part of traffic states or travel time to be model inputs or outputs. Moreover, comparing with ground-truth traffic phenomena is an effective way to validate microscopic (car-following and lane-changing) models from a macroscopic perspective.

• Demonstrating driving behaviors or traffic phenomena. Through graphically and statistically demonstrating and analyzing time-space vehicle trajectories, typical driving behaviors or traffic phenomena could be obtained and correspondingly modeled. The usage is usually adopted in car-following and lane-changing modeling and behavior analysis, as well as traffic phenomena illustration before modeling.

• Being analysis samples. The studies in the subjects, such as *macroscopic traffic phe-*

nomena estimation and analysis, *traffic flow model calibration*, and *vehicle trajectory data cleaning*, need samples to demonstrate and apply their proposed methods. The high-fidelity NGSIM trajectory data could satisfy the demands and play the role.

• Building simulation environments or testbeds. For the studies related to new technology that has not been practically applied, such as connected vehicles, VANET, safety and green driving strategies, the tests are usually conducted in simulation scenarios. Therefore, typical and representative trajectories or platoons can be selected from the real traffic data, and used to build simulation environments as realistic as possible.

6. Limitations

For current NGSIM trajectory datasets, the following limitations might exist.

• **Limitation-1**: The time-space scope of the NGSIM trajectories is limited. For the freeway (US-101 and I-80) datasets, the time coverages are 45 min, and the space coverages are approximately 640 m and 500 m, respectively. For the arterial (Peachtree and Lankershim) datasets, the time coverages are 30 min, and the space coverages are approximately 640 m (five intersections) and 488 m (four intersections), respectively.

Such time-space coverages are too short, due to the facts that an oscillation could propagate for 5-10 kilometers (Treiber and Kesting, 2012), and that a rush hour of traffic usually lasts for more than one hours. The shortage may be the most serious issue limiting the applications of the NGSIM datasets, as reported in Herrera and Bayen (2010); Treiber and Kesting (2012); Bifulco *et al.* (2013); Blandin *et al.* (2013); Wu and Liu (2014); Li *et al.* (2014); He *et al.* (2015); Coifman (2015); Oh and Yeo (2015); Jiang *et al.* (2015). In particular, more high-fidelity data, which contain not only car-following behavior but also a large amount of lane-changing behavior, are very meaningful for data-driven modeling of car-following behavior, or lane-changing behavior that has not been seen yet.

• **Limitation-2**: The traffic conditions contained in the NGSIM datasets are limited. The traffic conditions contained in the freeway datasets are congested traffic conditions with oscillating features, except some pieces of high flow traffic exhibited by the exiting traffic in the US-101 dataset and the HOV-lane traffic in the I-80 dataset. The limitation of lack of high-flow traffic are reported by the literature, such as Chiabaut *et al.* (2010); Li *et al.* (2013); Jin *et al.* (2015); Balal *et al.* (2016).

In contrast, the traffic contained in the arterial datasets is unsaturated without residual queues (Sun and Ban, 2013; Argote-Cabañero *et al.*, 2015). Therefore, some literature has to evaluate the model performance under saturated conditions in a simulation environment instead of using the arterial datasets. Lack of saturated conditions is also a main reason that the studies based on the arterial datasets are less than those based on the freeway datasets[8].

• **Limitation-3**: The variety of data-collection roads is limited. The NGSIM datasets were collected only from freeways and arterials in U.S., making impossible to investigate the traffic on other level or type of roads or on the roads in other countries . For example, urban freeways in some countries, such as China, play a leading role in their road networks. Compared with freeways, urban freeways have different geometry features, such as denser and shorter ramps, and lower speed limits (60 km/h-80 km/h). However, due to lack of high-fidelity datasets such as the NGSIM datasets, some unique traffic phenomena (e.g., the slowly moving but barely jammed congestion, high-speed car-following behavior) and their formation mechanisms are still unclear (Guan and He, 2008; He *et al.*, 2017).

Moreover, it is a question that to what extent the traffic phenomena or characteristics founded in the NGSIM datasets could be treated as a universal law; in other words, to what extent the NGSIM datasets could represent general traffic. Probably, only when more high-fidelity data are collected on various roads and in different locations, the question could be properly answered.

• **Limitation-4**: The variety of traffic components is limited. In general, vehicles can be classified in three types, i.e., motorcycles, passenger vehicles, and trucks, which are also distinguished by the NGSIM datasets. It is known that trucks with larger sizes impact on traffic more greatly than motorcycles and passenger vehicles. However, the NGISM datasets contain very limited trucks; for example, in the I-80 dataset, more than 80% vehicles are passenger vehicles with lengths smaller than 7 m (Coifman, 2015). Therefore, it is difficult to explicitly study on trucks' behavior and impact by using the NGSIM datasets.

In addition, the traffic at intersection are more complicated due to the mixture of pedestrians, bicycles (sometimes), and vehicles (Zeng *et al.*, 2014; Lu *et al.*, 2016). Unfortunately, the NGSIM datasets only contain vehicles, making the studies on the intersection traffic limited.

• **Limitation-5**: Relevant data are lacking, such as emissions, demographic and psychological information. Due to the way of collecting the NGSIM trajectories (i.e., cameras), other data related are still lacking, such as emissions, demographic and psychological information. Although simultaneously collecting those data in a large time-space scope is more like an "extravagant hope", there is no doubt that trajectory data with abundant relevant data will greatly widen and deepen our transportation research (and more papers will appear in TR-D and TR-F). It may be quite difficult to collect such data under natural conditions, and controlled experiments may be needed. However, it is obvious that large-scale controlled

[8] Another reason may be that more high-resolution traffic data at intersection are available (Liu *et al.*, 2009), because collecting traffic data at intersection is easier that collecting freeway traffic data.

experiments will be rather expensive.

7. Outlook

Considering the limitations of the NGSIM datasets, the following outlooks are given. Note that Outlooks 1-5 are correspondingly given to mitigate Limitations 1-5, and thus the specific explanations are omitted.

• **Outlook-1**: High-fidelity trajectory data with larger time-space scopes are expected.

• **Outlook-2**: High-fidelity trajectory data with various traffic conditions are expected.

• **Outlook-3**: High-fidelity trajectory data collected on various roads are expected.

• **Outlook-4**: High-fidelity trajectory data containing various traffic components are expected.

• **Outlook-5**: High-fidelity trajectory data with various aspects of related information are expected.

• **Outlook-6**: High-fidelity trajectory data with the applications of new technology are expected. It can be expected that connected vehicles (Uhlemann, 2015) and autonomous vehicles (SAE International, 2016) will largely appear on roads in short future, with the rapidly development of connected and autonomous vehicle technology. Therefore, it will be meaningful if such mixed traffic data could be collected.

Different from ten years ago, to collect high-fidelity vehicle trajectory data, it may be not necessary to place cameras in a flying helicopter (Hoogendoorn *et al.*, 2003) or on the top of a roadside high building (U.S. Federal Highway Administration, 2006) any more. Instead, a rotary wing Unmanned Aerial Vehicle (UAV, or called drone) with tethered power supply is recommended to carry high definition cameras to collect the high-fidelity traffic data with a large time-space scope; see Figure 3 for an illustration. In addition to the advantages of the traditional UAV, such as low cost, high flight altitude, and relatively good stability (Barmpounakis *et al.*, 2016), the tethered UAV is able to stay in sky with much longer time, such as hours or even days, because the tether cable can sustainably supply power to the UAV. The advantage of long flight time is significant to collect traffic data with a large time scope. Nevertheless, UAVs are vulnerable to adverse weather, and advanced technology is needed to process their collected imagines (Xu *et al.*, 2016a,b). Those disadvantages may limit their applications in traffic data collection.

8. Summary

This paper reviews the literature conducted based on the well-known high-fidelity NGSIM trajectory datasets including two freeway-traffic (i.e., US-101 and I-80) datasets and two arterial-traffic datasets (i.e., Peachtree and Lankershim). Due to the existence of a large

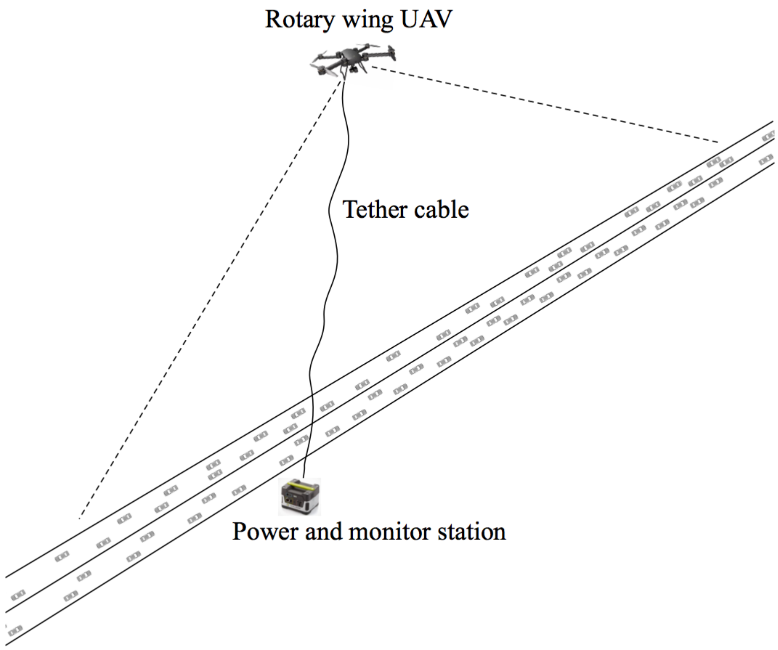

Figure 3: A schematic diagram of traffic data collection using a rotary wing UAV with tethered power supply.

amount of relevant literature, only the papers published in the journal series of Transportation Research are introduced, and all these studies are classified into six subjects, i.e., *microscopic traffic flow analysis and modeling*, *mesoscopic and macroscopic traffic flow modeling*, *traffic-related estimation and prediction*, *traffic flow model calibration*, *vehicle trajectory data cleaning*, and *VANET-related studies*.

It is found that the main usages of the NGSIM datasets in the literature are calibrating or training traffic flow models, validating models, demonstrating driving behaviors or traffic phenomena, being analysis samples, and building simulation environments or testbeds.

Five limitations (see the highlights in Section 6) of the NGSIM datasets are pointed out, and six outlooks (see the highlights in Section 7) are accordingly given to shed light on future traffic flow data collection and data-based studies. Moreover, a rotary wing UAV with tethered power supply is recommended to carry out the task of traffic data collection in future, due to its advantages such as low cost, high flight altitude, good stability, and long flight time.

References

U.S. Federal Highway Administration. Next Generation Simulation Program (NGSIM). *http://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm* **2006**.

Lijun, S.; Yafeng, Y. Discovering themes and trends in transportation research using topic modeling. *Transportation Research Part C: Emerging Technologies* **2017**, *77*, 49–66.

Tordeux, A.; Lassarre, S.; Roussignol, M. An adaptive time gap car-following model. *Transportation Research Part B: Methodological* **2010**, *44*, 1115–1131.

Koutsopoulos, H.N.; Farah, H. Latent class model for car following behavior. *Transportation Research Part B: Methodological* **2012**, *46*, 563–578.

Chen, D.; Laval, J.; Zheng, Z.; Ahn, S. A behavioral car-following model that captures traffic oscillations. *Transportation Research Part B: Methodological* **2012**, *46*, 744–761.

Laval, J.A.; Toth, C.S.; Zhou, Y. A parsimonious model for the formation of oscillations in car-following models. *Transportation Research Part B: Methodological* **2014**, *70*, 228–238.

Tian, J.; Treiber, M.; Ma, S.; Jia, B.; Zhang, W. Microscopic driving theory with oscillatory congested states: Model and empirical verification. *Transportation Research Part B: Methodological* **2015**, *71*, 138–157.

Wang, W.; Zhang, W.; Guo, H.; Bubb, H.; Ikeuchi, K. A safety-based approaching behavioural model with various driving characteristics. *Transportation Research Part C: Emerging Technologies* **2011**, *19*, 1202–1214.

Przybyla, J.; Taylor, J.; Jupe, J.; Zhou, X. Estimating risk effects of driving distraction: A dynamic errorable car-following model. *Transportation Research Part C: Emerging Technologies* **2015**, *50*, 117–129.

Hamdar, S.H.; Mahmassani, H.S.; Treiber, M. From behavioral psychology to acceleration modeling: Calibration, validation, and exploration of drivers' cognitive and safety parameters in a risk-taking environment. *Transportation Research Part B: Methodological* **2015**, *78*, 32–53.

Zheng, J.; Suzuki, K.; Fujita, M. Car-following behavior with instantaneous driver-vehicle reaction delay: A neural-network-based methodology. *Transportation Research Part C: Emerging Technologies* **2013**, *36*, 339–351.

He, Z.; Zheng, L.; Guan, W. A simple nonparametric car-following model driven by field data. *Transportation Research Part B: Methodological* **2015**, *80*, 185–201.

Hao, H.; Ma, W.; Xu, H. A fuzzy logic-based multi-agent car-following model. *Transportation Research Part C: Emerging Technologies* **2016**, *69*, 477–496.

Chiabaut, N.; Leclercq, L.; Buisson, C. From heterogeneous drivers to macroscopic patterns in congestion. *Transportation Research Part B: Methodological* **2010**, *44*, 299–308.

Chen, D.; Laval, J.A.; Ahn, S.; Zheng, Z. Microscopic traffic hysteresis in traffic oscillations: A behavioral perspective. *Transportation Research Part B: Methodological* **2012**, *46*, 1440–1453.

Chen, D.; Ahn, S.; Laval, J.; Zheng, Z. On the periodicity of traffic oscillations and capacity drop: The role of driver characteristics. *Transportation Research Part B: Methodological* **2014**, *59*, 117–136.

Wei, D.; Liu, H. Analysis of asymmetric driving behavior using a self-learning approach. *Transportation Research Part B: Methodological* **2013**, *47*, 1–14.

Li, L.; Chen, X.; Li, Z. Asymmetric stochastic Tau Theory in car-following. *Transportation Research Part F: Traffic Psychology and Behaviour* **2013**, *18*, 21–33.

Taylor, J.; Zhou, X.; Rouphail, N.M.; Porter, R.J. Method for investigating intradriver heterogeneity using vehicle trajectory data: A Dynamic Time Warping approach. *Transportation Research Part B: Methodological* **2015**, *73*, 59–80.

Hamdar, S.H.; Qin, L.; Talebpour, A. Weather and road geometry impact on longitudinal driving behavior: Exploratory analysis using an empirically supported acceleration modeling framework. *Transportation Research Part C: Emerging Technologies* **2016**, *67*, 193–213.

Laval, J.A.; Leclercq, L. Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model. *Transportation Research Part B: Methodological* **2008**, *42*, 511–522.

Zheng, Z.; Ahn, S.; Chen, D.; Laval, J. The effects of lane-changing on the immediate follower: Anticipation, relaxation, and change in driver characteristics. *Transportation Research Part C: Emerging Technologies* **2013**, *26*, 367–379.

Talebpour, A.; Mahmassani, H.S.; Hamdar, S.H. Modeling lane-changing behavior in a connected environment: A game theory approach. *Transportation Research Part C: Emerging Technologies* **2015**, *59*, 216–232.

Balal, E.; Cheu, R.L.; Sarkodie-Gyan, T. A binary decision model for discretionary lane changing move based on fuzzy inference system. *Transportation Research Part C: Emerging Technologies* **2016**, *67*, 47–61.

Yang, H.; Jin, W.L. A control theoretic formulation of green driving strategies based on inter-vehicle communications. *Transportation Research Part C: Emerging Technologies* **2014**, *41*, 48–60.

Gong, S.; Shen, J.; Du, L. Constrained optimization and distributed computation based car following control of a connected and autonomous vehicle platoon. *Transportation Research Part B: Methodological* **2016**, *94*, 314–334.

Tak, S.; Woo, S.; Yeo, H. Study on the framework of hybrid collision warning system using loop detectors and vehicle information. *Transportation Research Part C: Emerging Technologies* **2016**, *73*, 202–218.

Ahmed, K. Modeling drivers' acceleration and lane changing behaviors. PhD thesis, MIT, 1999.

Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **1974**, *19*, 716–723.

Newell, G. A simplified car-following theory: A lower order model. *Transportation Research Part B: Methodological* **2002**, *36*, 195–205.

Kerner, B. *Introduction to Modern Traffic Flow Theory and Control: The Long Road to Three-Phase Traffic Theory*; Springer, 2009.

Hamdar, S.H.; Treiber, M.; Mahmassani, H.S. Calibration of a Stochastic Car-Following Model Using Trajectory Data: Exploration and Model Properties. 88th Annual Meeting Transportation Research Board; , 2009.

Lee, D. A theory of visual control of braking based on information about time-to-collision. *Perception* **1976**, *5*, 437–459.

Kaneman, D.; Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica* **1979**, *47*, 263–292.

Chiu, Y.C.; Zhou, L.; Song, H. Development and calibration of the Anisotropic Mesoscopic Simulation model for uninterrupted flow facilities. *Transportation Research Part B: Methodological* **2010**, *44*, 152–174.

Daganzo, C.F. Requiem for second-order fluid approximations of traffic flow. *Transportation Research Part B* **1995**, *29*, 277–286.

Piccoli, B.; Han, K.; Friesz, T.L.; Yao, T.; Tang, J. Second-order models and traffic data from mobile sensors. *Transportation Research Part C: Emerging Technologies* **2015**, *52*, 32–56.

Qian, Z.; Li, J.; Li, X.; Zhang, M.; Wang, H. Modeling heterogeneous traffic flow: A pragmatic approach. *Transportation Research Part B: Methodological* **2017**, *99*, 183–204.

Jin, W.L. A kinematic wave theory of lane-changing traffic flow. *Transportation Research Part B: Methodological* **2010**, *44*, 1001–1021.

Jin, W.L. A multi-commodity Lighthill-Whitham-Richards model of lane-changing traffic flow. *Transportation Research Part B: Methodological* **2013**, *57*, 361–377.

Coifman, B. Empirical flow-density and speed-spacing relationships: Evidence of vehicle length dependency. *Transportation Research Part B: Methodological* **2015**, *78*, 54–65.

Siqueira, A.F.; Peixoto, C.J.T.; Wu, C.; Qian, W.L. Effect of stochastic transition in the fundamental diagram of traffic flow. *Transportation Research Part B: Methodological* **2016**, *87*, 1–13.

Jabari, S.E.; Zheng, J.; Liu, H.X. A probabilistic stationary speed-density relation based on Newell's simplified car-following model. *Transportation Research Part B: Methodological* **2014**, *68*, 205–223.

Wu, L.; Coifman, B. Improved vehicle classification from dual-loop detectors in congested traffic. *Transportation Research Part C: Emerging Technologies* **2014**, *46*, 222–234.

Laval, J.A. Hysteresis in traffic flow revisited: An improved measurement method. *Transportation Research Part B: Methodological* **2011**, *45*, 385–391.

Ahn, S.; Vadlamani, S.; Laval, J. A method to account for non-steady state conditions in measuring traffic hysteresis. *Transportation Research Part C: Emerging Technologies* **2013**, *34*, 138–147.

Zheng, Z.; Ahn, S.; Chen, D.; Laval, J. Applications of wavelet transform for analysis of freeway traffic: Bottlenecks, transient traffic, and traffic oscillations. *Transportation Research Part B: Methodological* **2011**, *45*, 372–384.

Zheng, Z.; Ahn, S.; Chen, D.; Laval, J. Freeway traffic oscillations: Microscopic analysis of formations and propagations using Wavelet Transform. *Transportation Research Part B: Methodological* **2011**, *45*, 1378–1388.

Blandin, S.; Argote, J.; Bayen, A.M.; Work, D.B. Phase transition model of non-stationary traffic flow: Definition, properties and solution method. *Transportation Research Part B: Methodological* **2013**, *52*, 31–55.

Oh, S.; Yeo, H. Impact of stop-and-go waves and lane changes on discharge rate in recovery flow. *Transportation Research Part B: Methodological* **2015**, *77*, 88–102.

Li, X.; Cui, J.; An, S.; Parsafard, M. Stop-and-go traffic analysis: Theoretical properties, environmental impacts and oscillation mitigation. *Transportation Research Part B: Methodological* **2014**, *70*, 319–339.

Herrera, J.C.; Bayen, A.M. Incorporation of Lagrangian measurements in freeway traffic state estimation. *Transportation Research Part B: Methodological* **2010**, *44*, 460–481.

Deng, W.; Lei, H.; Zhou, X. Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach. *Transportation Research Part B: Methodological* **2013**, *57*, 132–157.

Bucknell, C.; Herrera, J.C. A trade-off analysis between penetration rate and sampling frequency of mobile sensors in traffic state estimation. *Transportation Research Part C: Emerging Technologies* **2014**, *46*, 132–150.

Argote-Cabañero, J.; Christofa, E.; Skabardonis, A. Connected vehicle penetration rate for estimation of arterial measures of effectiveness. *Transportation Research Part C: Emerging Technologies* **2015**, *60*, 298–312.

Ramezani, M.; Geroliminis, N. On the estimation of arterial route travel time distribution with Markov chains. *Transportation Research Part B: Methodological* **2012**, *46*, 1576–1590.

Feng, Y.; Hourdos, J.; Davis, G.A. Probe vehicle based real-time traffic monitoring on urban roadways. *Transportation Research Part C: Emerging Technologies* **2014**, *40*, 160–178.

Qi, H.; Wang, D.; Chen, P.; Bie, Y. Simulation of departure flow profile at stop lines for signal approach spillover. *Transportation Research Part C: Emerging Technologies* **2013**, *33*, 88–106.

Srivastava, A.; Jin, W.L.; Lebacque, J.P. A modified Cell Transmission Model with realistic queue discharge features at signalized intersections. *Transportation Research Part B: Methodological* **2015**, *81*, 302–315.

Sun, Z.; Ban, X.J. Vehicle trajectory reconstruction for signalized intersections using mobile traffic sensors. *Transportation Research Part C: Emerging Technologies* **2013**, *36*, 2680–283.

Hao, P.; Sun, Z.; Ban, X.J.; Guo, D.; Ji, Q. Vehicle index estimation for signalized intersections using sample travel times. *Transportation Research Part C: Emerging Technologies* **2013**, *36*, 513–529.

Hao, P.; Ban, X.J.; Guo, D.; Ji, Q. Cycle-by-cycle intersection queue length distribution estimation using sample travel times. *Transportation Research Part B: Methodological* **2014**, *68*, 185–204.

Sun, Z.; Zan, B.; Ban, X.J.; Gruteser, M. Privacy protection method for fine-grained urban traffic modeling using mobile sensors. *Transportation Research Part B: Methodological* **2013**, *56*, 50–69.

Yang, K.; Guler, S.I.; Menendez, M. Isolated intersection control for various levels of vehicle technology: Conventional, connected, and automated vehicles. *Transportation Research Part C: Emerging Technologies* **2016**, *72*, 109–129.

Lee, S.; Wong, S.C.; Li, Y.C. Real-time estimation of lane-based queue lengths at isolated signalized junctions. *Transportation Research Part C: Emerging Technologies* **2015**, *56*, 1–17.

Lee, S.; Wong, S. Group-based approach to predictive delay model based on incremental queue accumulations for adaptive traffic control systems. *Transportation Research Part B: Methodological* **2017**, *98*, 1–20.

Edie, L. Discussion of Traffic Stream Measurements and Definitions. 2nd International Symposium on the Theory of Traffic Flow, 1963, pp. 139–154.

Li, X.; Wang, X.; Ouyang, Y. Characterization of traffic oscillation propagation under nonlinear car-following laws. *Transportation Research Part B: Methodological* **2012**, *46*, 409–423.

Newell, G. A simplified theory of kinematic waves in highway traffic I: General theory. II: Queuing at freeway bottlenecks. III: Multi-destination flows. *Transportation Research Part B* **1993**, *27*, 281–313.

Daganzo, C.F. A variational formulation of kinematic waves: Basic theory and complex boundary conditions. *Transportation Research Part B: Methodological* **2005**, *39*, 187–196.

Ilgin Guler, S.a.; Menendez, M.a.; Meier, L.a. Using connected vehicle technology to improve the efficiency of intersections. *Transportation Research Part C: Emerging Technologies* **2014**, *46*, 121–131.

Li, X.; Wang, X.; Ouyang, Y. Prediction and field validation of traffic oscillation propagation under nonlinear car-following laws. *Transportation Research Part B: Methodological* **2012**, *46*, 409–423.

Rhoades, C.; Wang, X.; Ouyang, Y. Calibration of nonlinear car-following laws for traffic oscillation prediction. *Transportation Research Part C: Emerging Technologies* **2016**, *69*, 328–342.

Kim, I.; Kim, T.; Sohn, K. Identifying driver heterogeneity in car-following based on a random coefficient model. *Transportation Research Part C: Emerging Technologies* **2013**, *36*, 34–44.

Vieira da Rocha, T.; Leclercq, L.; Montanino, M.; Parzani, C.; Punzo, V.; Ciuffo, B.; Villegas, D. Does traffic-related calibration of car-following models provide accurate estimations of vehicle emissions? *Transportation Research Part D: Transport and Environment* **2015**, *34*, 267–280.

Li, L.; Chen, X.M.; Zhang, L. A global optimization algorithm for trajectory data based car-following model calibration. *Transportation Research Part C: Emerging Technologies* **2016**, *68*, 311–332.

Durrani, U.; Lee, C.; Maoh, H. Calibrating the Wiedemann's vehicle-following model using mixed vehicle-pair interactions. *Transportation Research Part C: Emerging Technologies* **2016**, *67*, 227–242.

Zhong, R.X.; Fu, K.Y.; Sumalee, A.; Ngoduy, D.; Lam, W.H.K. A cross-entropy method and probabilistic sensitivity analysis framework for calibrating microscopic traffic models. *Transportation Research Part C: Emerging Technologies* **2016**, *63*, 147–169.

Sopasakis, A.; Katsoulakis, M.A. Information metrics for improved traffic model fidelity through sensitivity analysis and data assimilation. *Transportation Research Part B: Methodological* **2016**, *86*, 1–18.

Chandler, R.E.; Herman, R.; Montroll, E.W. Traffic Dynamics: Studies in Car Following. *Operations Research* **1958**, *6*, 165–184.

Aghabayk, K.; Sarvi, M.; Young, W.; Kautzsch, L. A novel methodology for evolutionary calibration of VISSIM by multi-threading. 36th Australasian Transport Research Forum; , 2013.

Punzo, V.; Borzacchiello, M.T.; Ciuffo, B. On the assessment of vehicle trajectory data accuracy and application to the Next Generation SIMulation (NGSIM) program data. *Transportation Research Part C: Emerging Technologies* **2011**, *19*, 1243–1262.

Zheng, Z.; Washington, S. On selecting an optimal wavelet for detecting singularities in traffic and vehicular data. *Transportation Research Part C: Emerging Technologies* **2012**, *25*, 18–33.

Montanino, M.; Punzo, V. Trajectory data reconstruction and simulation-based validation against macroscopic traffic patterns. *Transportation Research Part B: Methodological* **2015**, *80*, 82–106.

Fard, M.R.; Mohaymany, A.S.; Shahri, M. A new methodology for vehicle trajectory reconstruction based on wavelet analysis. *Transportation Research Part C* **2017**, *74*, 150–167.

Zheng, Z.; Su, D. Traffic state estimation through compressed sensing and Markov random field. *Transportation Research Part B: Methodological* **2016**, *91*, 525–554.

Baiocchi, A. Analysis of timer-based message dissemination protocols for inter-vehicle communications. *Transportation Research Part B: Methodological* **2016**, *90*, 105–134.

Du, L.; Gong, S.; Wang, L.; Li, X.Y. Information-traffic coupled cell transmission model for information spreading dynamics over vehicular ad hoc network on road segments. *Transportation Research Part C: Emerging Technologies* **2016**, *73*, 30–48.

Treiber, M.; Kesting, A. Validation of traffic flow models with respect to the spatiotemporal evolution of congested traffic patterns. *Transportation Research Part C: Emerging Technologies* **2012**, *21*, 31–41.

Bifulco, G.N.; Pariota, L.; Brackstione, M.; Mcdonald, M. Driving behaviour models enabling the simulation of Advanced Driving Assistance Systems: Revisiting the Action Point paradigm. *Transportation Research Part C: Emerging Technologies* **2013**, *36*, 352–366.

Wu, X.; Liu, H.X. Using high-resolution event-based data for traffic modeling and control: An overview. *Transportation Research Part C: Emerging Technologies* **2014**, *42*, 28–43.

Jiang, R.; Hu, M.B.; Zhang, H.M.; Gao, Z.Y.; Jia, B.; Wu, Q.S. On some experimental features of car-following behavior and how to model them. *Transportation Research Part B: Methodological* **2015**, *80*, 338–354.

Jin, C.J.; Wang, W.; Jiang, R.; Zhang, H.M.; Wang, H.; Hu, M.B. Understanding the structure of hyper-congested traffic from empirical and experimental evidences. *Transportation Research Part C: Emerging Technologies* **2015**, *60*, 324–338.

Liu, H.X.; Wu, X.; Ma, W.; Hu, H. Real-time queue length estimation for congested signalized intersections. *Transportation Research Part C: Emerging Technologies* **2009**, *17*, 412–427.

Guan, W.; He, S. Statistical features of traffic flow on urban freeways. *Physica A: Statistical Mechanics and its Applications* **2008**, *387*, 944–954.

He, Z.; Zheng, L.; Chen, P.; Guan, W. Mapping to Cells: A Simple Method to Extract Traffic Dynamics from Probe Vehicle Data. *Computer-Aided Civil and Infrastructure Engineering* **2017**, *32*, 252–267.

Zeng, W.; Chen, P.; Nakamura, H.; Iryo-Asano, M. Application of social force model to pedestrian behavior analysis at signalized crosswalk. *Transportation Research Part C: Emerging Technologies* **2014**, *40*, 143–159.

Lu, L.; Ren, G.; Wang, W.; Chan, C.Y.; Wang, J. A cellular automaton simulation model for pedestrian and vehicle interaction behaviors at unsignalized mid-block crosswalks. *Accident Analysis and Prevention* **2016**, *95*, 425–437.

Uhlemann, E. Introducing connected vehicles. *IEEE Vehicular Technology Magazine* **2015**, *10*.

SAE International. Automated driving levels of driving automation are defined in new SAE international standard J3016. *http://www.sae.org/misc/pdfs/automated_driving.pdf* **2016**.

Hoogendoorn, S.; Van Zuylen, H.; Schreuder, M.; Gorte, B.; Vosselman, G. Microscopic Traffic Data Collection by Remote Sensing. *Transportation Research Record: Journal of the Transportation Research Board* **2003**, *1855*, 121–128.

Barmpounakis, E.N.; Vlahogianni, E.I.; Golias, J.C. Unmanned Aerial Systems for Transportation Engineering: Current practice and future challenges. *International Journal of Transportation Science and Technology* **2016**, *5*, 111–122.

Xu, Y.; Yu, G.; Wu, X.; Wang, Y.; Ma, Y. An Enhanced Viola-Jones Vehicle Detection Method From Unmanned Aerial Vehicles Imagery. *IEEE Transactions on Intelligent Transportation Systems* **2016**, *18*, 1845–1856.

Xu, Y.; Yu, G.; Wang, Y.; Wu, X.; Ma, Y. A hybrid vehicle detection method based on viola-jones and HOG + SVM from UAV images. *Sensors* **2016**, *16*.