Developing Extended Trajectory Database for Heterogeneous Traffic like NGSIM Database

Raju, Narayana; Arkatkar, Shriniwas; Easa, Said; joshi, gaurang

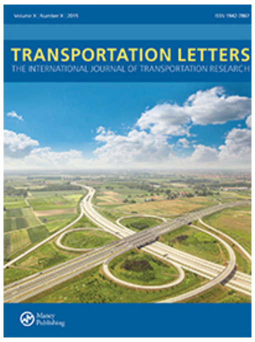
DOI
10.1080/19427867.2021.1908490

Publication date
2021

Document Version
Final published version

Published in
Transportation Letters: the international journal of transportation research

Citation (APA)
Raju, N., Arkatkar, S., Easa, S., & joshi, G. (2021). Developing Extended Trajectory Database for Heterogeneous Traffic like NGSIM Database. *Transportation Letters: the international journal of transportation research*, *14*(5), 555-564. https://doi.org/10.1080/19427867.2021.1908490



Taylor & Francis
Taylor & Francis Group

Transportation Letters
The International Journal of Transportation Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ytrl20

Developing extended trajectory database for heterogeneous traffic like NGSIM database

Narayana Raju, Shriniwas Arkatkar, Said Easa & Gaurang Joshi

TRANSPORTATION LETTERS
https://doi.org/10.1080/19427867.2021.1908490

Developing extended trajectory database for heterogeneous traffic like NGSIM database

Narayana Raju[a,b], Shriniwas Arkatkar[c], Said Easa[d] and Gaurang Joshi[e]

[a]Postdoctoral Researcher, Transportation and Planning, Delft University of Technology, Stevinweg, The Netherlands; [b]Former Research Scholar, Civil Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India; [c]Associate Professor, Civil Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India; [d]Professor, Civil Engineering Department, Ryerson University, Toronto, Canada; [e]Professor, Civil Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India

ABSTRACT
The present work introduced a framework of developing comprehensive extended vehicular trajectory data under heterogeneous non-lane-based traffic conditions like the NGSIM datasets in the United States. Due to the absence of automation and instrumentation, and even the lack of sensor deployment on roads in developing economies like India, it is even more challenging to study driver behavior. A new stitching-based algorithm was used for developing the extended trajectory database for three traffic-flow levels on a 535-m long section of an urban arterial. The algorithm was used to stitch the trajectory data over the segments such that the subject vehicle with continuous trajectory data points over the entire study stretch. The developed framework is a novel tool for establishing a trajectory dataset for mixed traffic, it should be of interest to researchers in developing and developed countries.

KEYWORDS
stitching logic; trajectory data; NGSIM; mixed traffic

Introduction

Driving behavior on a road section is one of the multifarious phenomena in traffic science. Simultaneously, various performance measures such as microscopic traffic characteristics, stability, safety in the traffic stream, and other efficacy studies of geometric elements are directly associated with driving behavior. Initially, researchers largely depended on the mathematical car-following models to model the driving behavior (Forbes et al. 1959; Gazis, Herman, and Potts 1959; Pipes 1953). Considering human intervention in driving behavior, later researchers such as Gipps (1981), Wiedemann (1974), and Newell (2002) framed different psychophysical car-following models. Besides, numerous other models were conceptualized, and driving behavior was modeled. Given the data constraints in sensing the drivers' retorting nature, the performance of the models was limited to a specific traffic scenario with the accentuated conceptual framework.

In parallel, researchers also performed various experimental trials over the test network to quantify driving behavior to capture the driving behavioral instincts. From these experiments, the notable outcome such as the drivers' reaction time was assessed (Johansson and Rumar 1971), which in turn played a massive role in the geometric design of highway elements, such as stopping, passing, and decision sight distances. Further, the advances in technology and the availability of scientifically high-end computational tools, such as simulation software and driving simulators, provided numerous opportunities to capture driving behavior. Traffic simulation tools allowed collecting different behavioral rationalities, such as car-following models (mostly psychophysical models), lane-changing logics, and gap-acceptance models. These tools were combined in simulation experiments as a package for modeling driving behavior and provided a great opportunity in mimicking the real-field conditions reasonably well, depending on the credibility of the models used.

Based on the preceding analytical and experimental tools, studies of many highways aspects are reported, including weaving analysis (Hidas 2005), safety performance (Sobhani, Young, and Sarvi 2013), transit operations (Toledo et al. 2010), adaptive cruise control (Kesting et al. 2008), and traffic emissions (Huang, Bird, and Bell 2009). With numerous valuable outcomes, traffic flow studies are also taken up to the next level. Again, at the same time, the studies resulted in highly customized options in simulation models, along with diversely rich plenteous microscopic field inputs that have allowed well-calibrated models or simulation process. However, capturing such microscopic interactions among vehicles from field conditions require the development of high-quality traffic databases that can capture vehicle retorts. This represented a substantial research gap in this direction that resulted in less confidence in simulation models' results. On the other hand, with the advent of using driving simulators, the receptiveness of drivers is well-probed for the coded conditions, including the use of mobile phones (Haque and Washington 2015), text messaging (Drews et al. 2009), mandatory lane changes (Ali, Zheng, and Haque 2018; Shirke et al. 2017), hazard perception (Underwood, Crundall, and Chapman 2011), and weather conditions (Kilpeläinen and Summala 2007). While such studies encapsulated the factors related to driving instincts based on laboratory experiments, the practical applicability of the outcomes in assessing real field driving behavior is uncertain.

Nonetheless, researchers have perceived the importance of capturing driving behavior from the traffic stream. Different variants of probed vehicles (mostly embeds a video camera and GPS) were used, and their performance was related to traffic stream characteristics. Some of the notable studies include traffic calming (Lee et al.

CONTACT Shriniwas Arkatkar sarkatkar@gmail.com Former Research Scholar, Civil Engineering Department, Sardar Vallabhbhai National Institute of Technology, Surat, India-

This article has been republished with minor changes. These changes do not impact the academic content of the article.

2013), travel time studies (Jenelius and Koutsopoulos 2013), performance characterization (Remias et al. 2013), and vehicle penetration (Feng et al. 2010). To a certain extent, using traffic stream data, probed-vehicle studies gaged traffic stream delay, platooning of vehicles, driving cycle, and following behavior (with multiple probed vehicles). However, probed-vehicle studies have not well defined the microscopic interactions among vehicles. Given the stochastic nature of driving behavior, representing traffic stream behavior with a single or a few vehicles can be doubtful. Keeping this in view, the study of driving behavior warrants high-quality trajectory data, where vehicular positions from the traffic stream must be tracked with the smallest possible update interval.

Assessing this research gap, the U.S. Federal Highway Administration (FHWA) as part of the next-generation simulation (NGSIM) project (FHWA 2007) developed a vehicle trajectory dataset using automated image processing tools at different locations for trap lengths of 400–600 m for different classes of roadway facilities. The NGSIM dataset serves as one of the prime sources in understating driving behavior under the lane-based homogeneous traffic conditions prevailing in the United States. On these lines, using micro-level data, numerous studies are reported across different parts of the world, including shock-wave analysis (Lu and Skabardonis 2007), traffic-flow assessment (Montanino and Punzo 2013), lane-changing behavior (Leclercq et al. 2007), traffic-oscillation analysis (Chen et al. 2012), simulation studies (Chiu, Zhou, and Song 2010; Kumar et al. 2020), and car-following behavior (Hao, Ma, and Xu 2016). Clearly, the development of NGSIM data emphasizes the importance of having micro-level high-quality vehicle trajectory data for understanding driving behavior in the best possible manner.

Unlike the lane-based homogeneous traffic conditions prevailing in developed economies, the traffic conditions prevailing in developing economies, such as India, are heterogeneous. The variety of vehicle classes on the roads coupled with ensuing weak lane-discipline results in complex interactions among vehicles due to the many possible combinations of longitudinal and lateral gaps, depending on static and dynamic characteristics. These complex interactions are challenging to understand. Hence, high-quality vehicle trajectory data are needed for determining simultaneous vehicle movements in both longitudinal and lateral directions under such mixed (heterogeneous) traffic conditions. Further, with the varied vehicle classes present on roads, even the well-established image processing tools failed to a large extent in developing the trajectory data at the desired level of accuracy. As a result, very few studies (Bharadwaj et al. 2016a, 2016b; Chunchu, Kalaga, and Seethepalli 2010; Kanagaraj et al. 2015; Munigety, Vicraman, and Mathew 2014; Raju et al. 2018, 2017) Clearly, the development of NGSIM data emphasizes the importance of having micro-level high-quality vehicle trajectory data for understanding driving behavior in the best possible manner.

Unlike the lane-based homogeneous traffic conditions prevailing in developed economies, the traffic conditions prevailing in developing economies, such as India, are heterogeneous. The variety of vehicle classes on the roads coupled with ensuing weak lane-discipline results in complex interactions among vehicles due to the many possible combinations of longitudinal and lateral gaps, depending on static and dynamic characteristics. These complex interactions are challenging to understand. Hence, high-quality vehicle trajectory data are needed for determining simultaneous vehicle movements in both longitudinal and lateral directions under such mixed (heterogeneous) traffic conditions. Further, with the varied vehicle classes present on roads, even the well-established image processing tools failed to a large extent in developing the trajectory data at the desired level of accuracy. As a result, very few studies (Kanagaraj et al. 2015) that used a length of 225 m. Considering this state of progress, it is quite convincing that the driving behavior prevailing under heterogeneous traffic conditions has not been explored much, unlike under homogeneous traffic conditions.

In the present work, an extended vehicle trajectory dataset was developed, analogous to the NGSIM dataset. According to NGSIM guidelines, a segment length not less than 500 m. This extended segment of half a kilometer may include four to five traps of 150–200 m. Then, the vehicles from different traps should be tracked over a space of 500 m for developing high-quality trajectory data. The next section describes the study area used for data collection. The following section presents the proposed methodology, including the conceptual framework, the concept of trajectory data development, challenges for extended trajectory dataset, trajectory stitching algorithm, smoothing trajectory data, and validation of the trajectory dataset. Investigation of driving behavior and discussion of results is then presented, followed by conclusions.

Study area

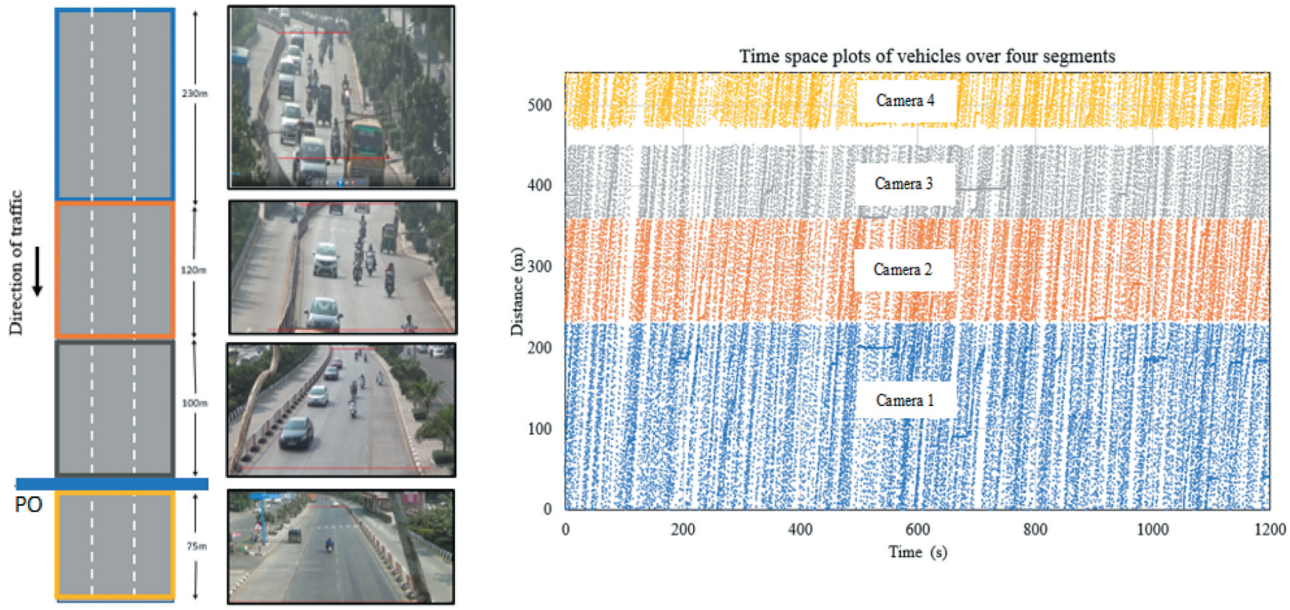
In this study, a mid-block section of urban arterial (road section web link) in Surat city, India, was selected for data collection. The section is about 535-m long and has three lanes in each direction of traffic flow (each 3.5-m wide), identified as median-side lane, middle lane, and shoulder-side lane, with a total width of 10.5 m. The study segment was selected so that it was away from any intersection and free from side-interference-producing activities. Within the study section, a pedestrian overpass (10-m wide) is located across the arterial at about 450 m from the start of the section with a vertical clearance of 7.5 m from the road surface. Considering the pedestrian overpass as a vantage point, four wide-angle digital cameras were installed and focused at four different continuous road segments with trap lengths of 230, 120, 100, and 75 m, respectively, as shown in Figure 1. This covered an extended study section of 535 m for developing a high-quality trajectory database using continuously captured data on vehicle movements over space and time under varying traffic flow conditions. The traffic flow comprised five categories of vehicles: motorized three-wheelers (MThW), motorized two-wheelers (MTW), cars, trucks, and light commercial vehicles (LCV). Note that the section length is more than half a kilometer and, therefore, consistent with the quality trajectory datasets recommended by NGSIM.

It can be noted that during of process of videography surveys for 12 consecutive hours, a wide range of traffic flow is observed over the study section from free flow to near capacity flow conditions. Macroscopic traffic characteristics, such as stream speed, flow, and density, are evaluated; thus, fundamental macroscopic diagrams are developed for the study section. Later traffic states are classified with the help of fundamental macroscopic diagrams.

Methodology

Conceptual framework

Vehicle trajectory data can be extremely a potent source in analyzing driving behavior at its best level. Further, with the limitation of trajectory data sources for heterogeneous traffic conditions, driving behavior for such traffic has not been explored comprehensively.



(a) Road segments (surveyed by 4 cameras) and pedestrian overpass (PO)

(b) Time-space diagram for vehicle movements over the study section

Figure 1. Study arterial segments surveyed by cameras and time space plots.

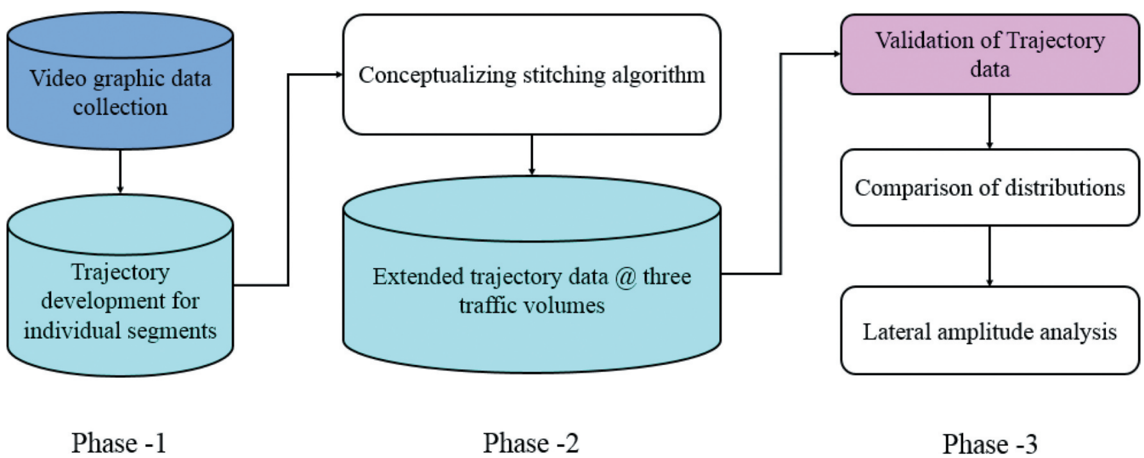


Figure 2. Research methodology adopted in the study.

Considering this research gap, a framework for establishing extended trajectory data for heterogeneous traffic conditions was developed. The framework involves three phases (see Figure 2). In the first phase, video graphic surveys were carried out to record vehicles' movements over the study section. In the second phase, a stitching-based algorithm was developed and used to establish an extended trajectory dataset like NGSIM. In the final phase, the established dataset was validated, followed by investigating driving behavior in terms of lateral amplitude.

Trajectory data development

Vehicle trajectory data are among the high-quality micro-level data sources that allow understanding of the interactions among vehicles. In the present study, vehicle trajectory data under prevailing mixed traffic conditions were developed. Initially, the preliminary analysis found that automated trajectory development tools are not efficient in trajectory development under non-lane-based mixed traffic conditions. Hence, a traffic data extractor (Vicraman et al. 2014), a semi-automated tool, was used to develop trajectory data. Further, with a semi-automated image processing tool, vehicle trajectory data were developed for the four individual segments with an update interval of 0.1 s. This was accomplished by tracking individual vehicles for each of the four segments, where the tracked vehicles were spotted with green dots. The vehicles were tracked manually using a computer mouse pointer concerning the vehicle's central position. Following this procedure, trajectory data at different flow levels were extracted for 20 min each, classified as Flow 1, Flow 2, and Flow 3, as shown in Table 1. A suitable and robust algorithm was then developed to stitch the trajectory data for each section to obtain extended trajectory data.

Table 1. Trajectory data analytics over the study section.

Flow condition	No. of vehicles tracked	Trajectory data duration (min)	Trafficcomposition (%) [a]	Traffic flow	Volume/capacity
Free flow	591	20 minutes	11.6, 51.1, 35.7, 0.8, 0.7	Flow 1	0.32
Moderate flow	891	20 minutes	12.8, 52.6, 29.5, 2.9, 1.7	Flow 2	0.43
Near capacity	1987	20 minutes	7.8, 57.6, 33.8, 0.22, 0.4	Flow 3	0.91

[a]Traffic composition: motorized three wheelers, motorized two wheelers, cars, trucks, and LCV.

Developing extended trajectory dataset

As previously mentioned, using the trajectory data extractor, vehicle trajectory data were developed over the four continuous road segments, for 20 min each at three flow levels for a total period of 60 min. However, to comprehensively evaluate driver behavior, a given subject vehicle must be traced in a continuous way over all four study sections. In the present case, with four cameras recording individually, the subject vehicle's identity is varied as the vehicle moves from one segment to another. As a result, the development of an extended trajectory is always articulated as a considerable research gap, especially under mixed traffic conditions.

To overcome this challenge, a logic is developed to track continuous vehicular movements. After close inspection of trajectory data, it can be noticed that for a given vehicle at a particular timestamp, the vehicular longitudinal position, lateral position, and category of the vehicle can be recorded by assigning a specific unique Id. In the present work, as trajectory data was developed individually for each of the four segments, the same vehicle was tracked with different Id numbers over different trap lengths. To address this issue, trajectory data from two continuous segments are selected as primary and secondary segments, in which traffic moves from the primary segment to the second segment in one direction of traffic flow. When a vehicle moves out from the primary segment, it can be projected that it will be detected in the second segment with a minor difference in the time stamp, having information such as longitudinal position, lateral position, and matching vehicle category. However, it might have been tracked with a different unique id in the second segment and can be well-visualized in Figure 3, in which when the red vehicle identity (number) will be changed once it exits the segment-1 to the other.

The present study's trajectory data were developed individually for each of the four segments; the same vehicle is tracked with different identification (ID) numbers over different trap lengths. To address this issue, trajectory data from two continuous segments were selected as primary and secondary segments, in which traffic moves from the primary segment to the second segment in one direction of traffic flow. When a vehicle moved out from the primary segment, it was detected in the second segment based on four specified thresholds related to the time stamp, longitudinal position, lateral position, and vehicle category. Since a vehicle might have been tracked with a different unique ID in the second segment, this ID would be replaced with the primary segment's ID.

The logic of the stitching algorithm is shown in Figure 4. In line with the insinuation, the algorithm substitutes vehicle ID in the second segment concerning the primary section. The algorithm's logic starts by assembling all trajectory endpoints

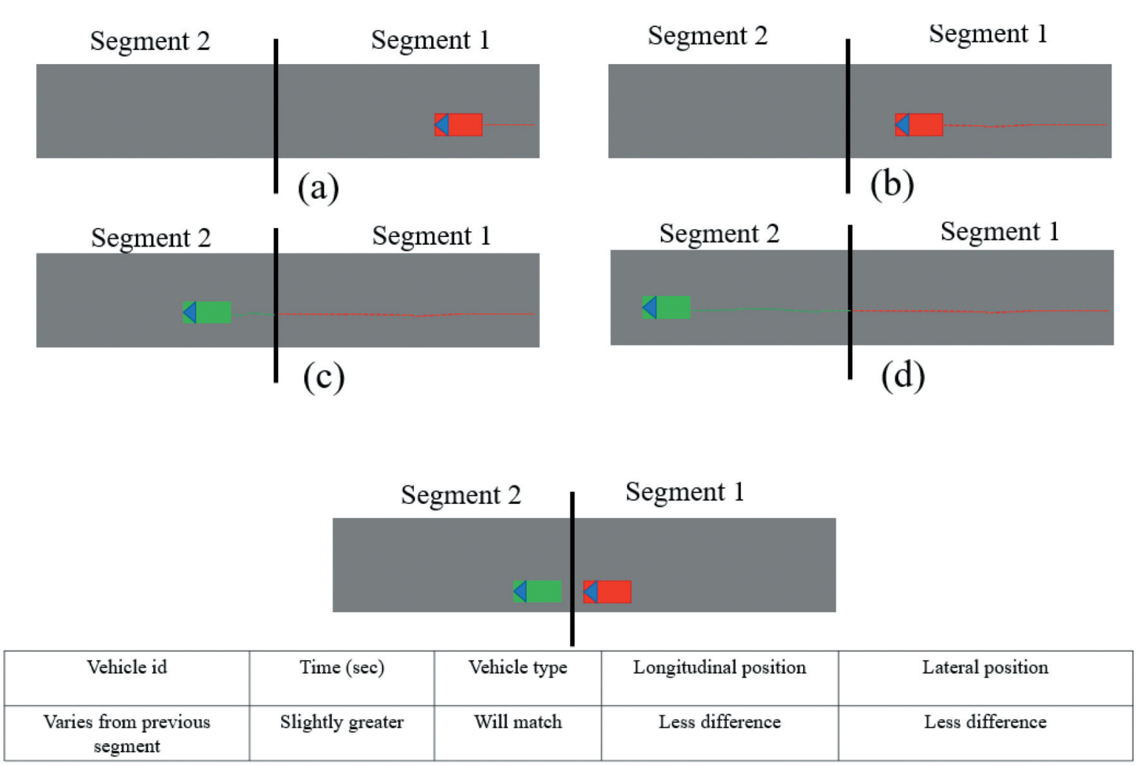


Figure 3. Trajectory data nature before switching over the two segments.

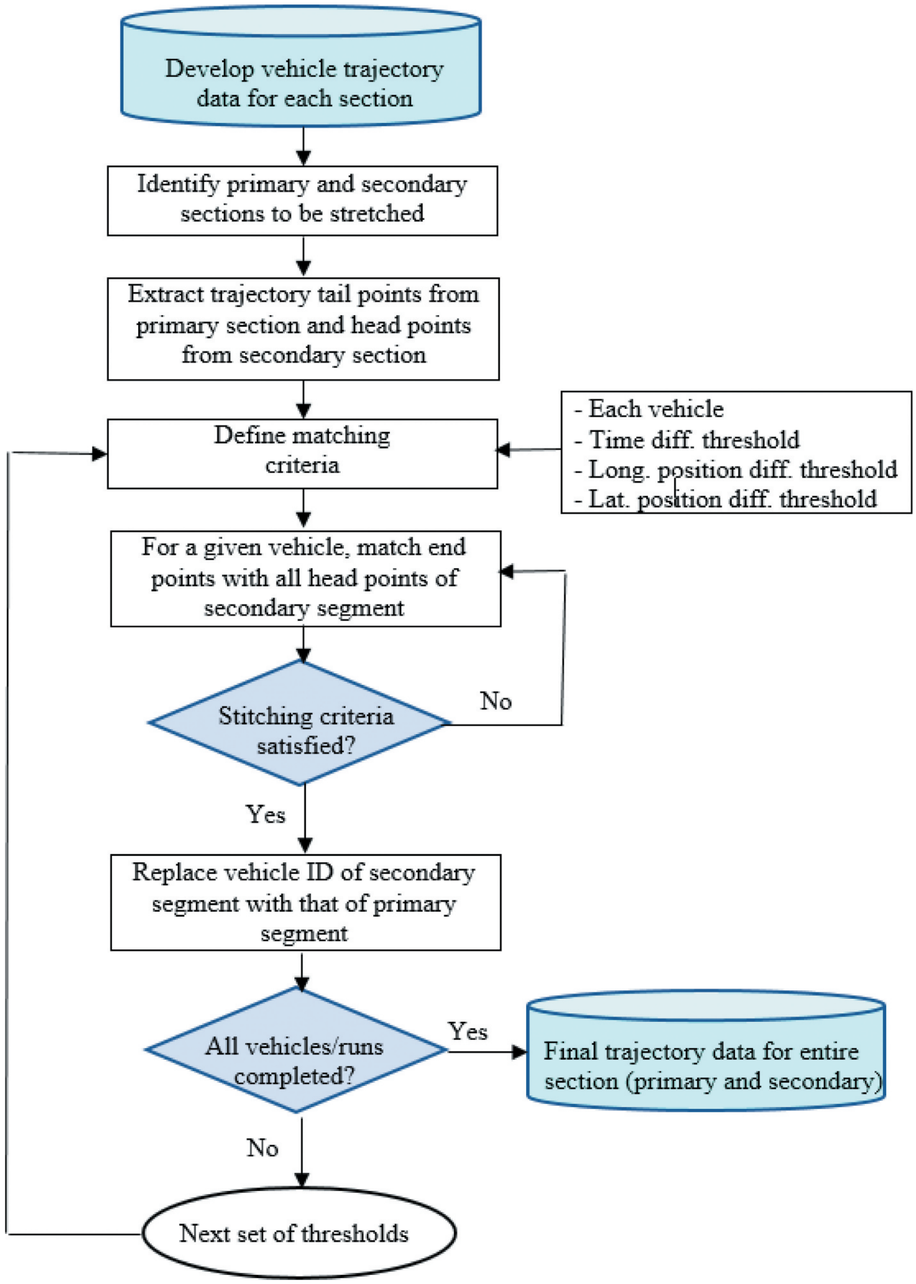


Figure 4. Algorithm coded in MATLAB for stitching trajectory data.

(tail points) from the primary section and the starting points (head points) from the secondary section. For matching the time stamp, longitudinal position, and lateral position, a set of thresholds are specified for different sequential runs, as shown in Table 2. Initially, the vehicle category was first matched, followed by the time difference, then longitudinal position difference, and lateral position difference. If a given head and tail parts satisfy the prescribed thresholds criteria, the second section's unique ID is replaced with the matched unique ID from the primary section. This recurrent procedure is repeated for all vehicles' tails in the primary section with the heads of all vehicles in the secondary section. To ensure good accuracy in stitching, the formulated logic's whole process was repeated for 10 runs with different progressively increased thresholds. This reduced the misclassification of vehicles while stitching a trajectory over a section of 535 m. In the present work, the vehicles are tracked very prudently over the space. Also, in the present work, lot of efforts are taken in deriving the positions of the vehicles while traversing from one frame to next frame both in lateral as well as longitudinal directions. Conventionally adopting the calibrated parameters and applying them uniformly for stitching empirical trajectory data can limit the data accuracy and is highly uncertain (particularly due to complexity in movement of vehicles under heterogeneous traffic conditions). Concurrently, to resolve this issue, the developed stitching algorithm has a good flexibility in adopting a different set of progressive thresholds (stage wise) over the several runs. In this direction, to understand the efficacy of the stitching logic, the section 1 trajectory data is divided into halves. In the second half, the Ids of the vehicles

Table 2. Thresholds used for stitching trajectory data over consecutive runs.

Runnumber	Vehicle category	Time diff. less than (s)	Longitudinal position diff. less than (m)	Lateral position diff. less than (m)	Remarks [a]
1	Should match	3	5	0.7	-
2	Should match	5	5	0.7	Time +2
3	Should match	7.5	5	0.7	Time +2.5
4	Should match	7.5	10	0.7	Long +5
5	Should match	9	10	0.7	Time +1.5
6	Should match	10	15	0.9	Time +1.5, long +5 lat+0.2
7	Should match	12	15	1.5	Time +2, lat+0.6
8	Should match	14	15	2.5	Time +2, lat+1
9	Should match	14	20	3	Long+5, lat+0.5
10	Should match	15	20	3.5	Time +1, lat+0.5

[a]Time = time difference, long = longitudinal distance difference, and lat = Lateral distance difference.

are replaced. Later the stitching algorithm is applied to stitch the trajectory data. In view of truly continuous nature of the trajectory data from the shorter section, the entire trajectory data is stitched with in the first run with 100% accuracy.

Smoothing trajectory data and amplifying continuity

It was anticipated that the pedestrian overpass would act as a blind spot, where the mounted video cameras would not capture trajectory data under the overpass. As previously mentioned, the 10-m wide overpass was located at 400 m from the starting between the third and fourth sections point (Figure 1). This can also be noted in the time–space plots shown in Figure 5. To address this constraint, the longitudinal-position difference threshold was increased by 15 m and the time-difference threshold by 5 s (based on average stream speed) for stitching the data between the third and fourth road segments. Further, to maintain continuity in the trajectory data over the entire road space, the missing data points were predicted using a high-order polynomial function that predicts longitudinal and lateral positions of individual vehicles as a function of time.

Let the longitudinal and lateral positions of individual vehicle at time t be represented, respectively, by $y(t)$ and $x(t)$, $t= t_1, t_2, \ldots, t_n$, where t_1 and t_n are time stamps 1 and n, respectively. Then, the fitted polynomials took the following form:

$$y(t) = \sum_{k=0}^{j} a_k(t)^k \tag{1}$$

$$x(t) = \sum_{k=0}^{j} a_k(t)^k \tag{2}$$

where $k$is the integer representing the degree of polynomial function, a_k is the coefficient term, and $t=$ time stamps.

Then, let t_m $(1 < m < n)$ be the time stamps where $y(t_m)$ and $x(t_m)$ are not recorded due to the blind spots caused by the pedestrian overpass. Using Equations 1 and 2, the longitudinal and lateral positions of the missing points, $y\left(t_m^k\right)$ and $x\left(t_m^k\right)$, are predicted by substituting $t= t_m$ in Equations 1 and 2.

To eliminate the noise in the trajectory dataset, smoothening technique such as moving average filter (Papailias and Thomakos 2015; Raju et al. 2017) was applied. Based on the stitching logic, an extended trajectory data for the entire 535-m section were developed. The developed time–space continuous plots using the proposed stitching algorithm over the entire study stretch is shown in Figure 5. As noted, the presence of vehicles over the longer segment even showing their imprints of the road carriageway in the extended time–space plots.

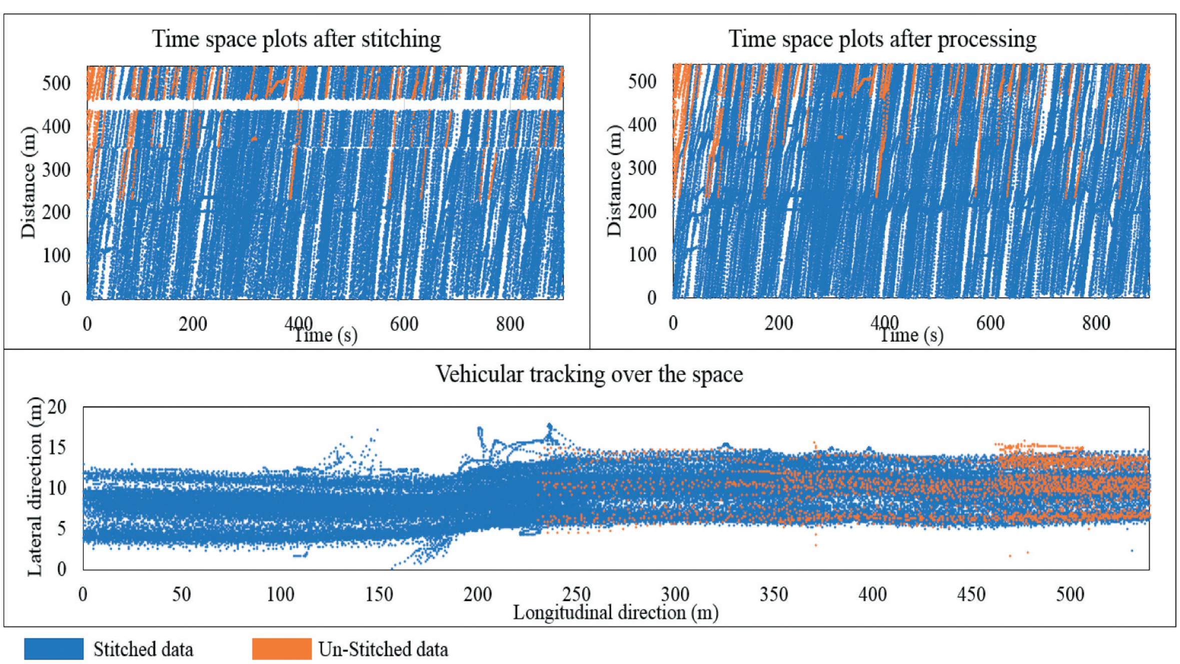


Figure 5. Time–space plots of the vehicles over the study section.

Validation

Two types of validation were performed: (1) validating the effectiveness of the stitching algorithm and (2) validating the extended trajectory database using field data. For the first validation, the algorithm was initially trained with some sample trajectory dataset of smaller trap lengths of 150–200 m (by dividing it into two to three parts that were initially developed data without any stitching logic). The results of this experiment showed an accuracy of about 95% on the trajectory database developed using the trajectory extractor. With the developed stitching logic, the stitching task was sequentially performed using the individually developed trajectory data for the four road segments (230, 120, 100, 75 m). It was found that the number of vehicles that are stitched accurately over the entire section was as follows: 584 out of 706 vehicles for Flow 1, 764 out of 891 vehicles for Flow 2, and 1621 out of 1987 vehicles for Flow 3, indicating an overall accuracy of about 80%. This was considered a high level of accuracy, given the non-lane-based vehicle movements of mixed traffic at the micro-level and the diverse driving behavior on Indian roads. Note that the developed trajectory data for the 535-m section length can be regarded as the first extended vehicle trajectory database for the Indian traffic condition context.

Forthe second validation, at the time of video-graphic surveys, 25-probe vehicle runs were carried over the study stretch. The predominant vehicle in this study (MTW) was selected as a probe vehicle, over which Velocity-Box (V-Box) instrument was mounted. The V-Box acted like the global positioning system (GPS) and traced the longitudinal position of MTW over the road space with an interval of 0.1 s. To test the developed trajectory dataset in the best possible manner in line with the literature (Durrani, Lee, and Maoh 2016), the second derivative of the position with respect to time (acceleration of the MTW vehicle) was compared and the probability density function established, as shown Figure 6. Further, to check the validity of the developed trajectory data, the Wilcoxon rank-sum test (Krishnamoorthy 2016) was performed to check the variability among the accelerations. From this test, a test statistic of 104 was observed, and at a 5% level of significance, the critical value was found to be 59, which is less than the critical test statistic 104, indicating no significant difference between the data. Based on this analysis, it is statistically concluded that the developed extended trajectory represents well the true vehicle movements over the road space under consideration.

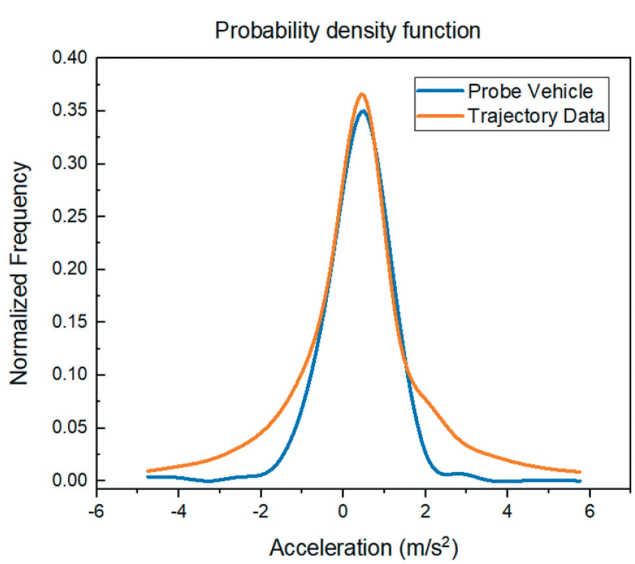


Figure 6. Probability density function of instant acceleration rates.

Analysis of driving behavior

Based on the developed extended trajectory data, driving behavior in terms of lateral amplitude (Raju, Arkatkar, and Joshi 2020) was evaluated. Lateral amplitude is the lateral weaving of vehicles, which is the difference in maximum and minimum lateral positions over the study section. Further, schematic representation that explains the lateral amplitude is shown in Figure 7(a). Mathematically, the lateral amplitude is given by

$$L_{amp} = L_{max} - L_{min} \tag{3}$$

$$L_{max} = \max[L(t)] t \in T \tag{4}$$

$$L_{min} = \min[L(t)] t \in T \tag{5}$$

where

L_{amp} is the lateral amplitude (m), L_{max} is the maximum lateral coordinate (m), L_{min} is the minimum lateral coordinate (m), $L(t)$= lateral position of the vehicles from trajectory data at time stamp t, and T is the time stamps of the vehicle

Based on the vehicle categories and their sizes, the data were segregated, in that MTW as one class, MThW and cars in one class, and bus, truck, and LCV in one class. Then, the lateral amplitude was evaluated based on vehicle class at all three flow levels. The distributions of the lateral amplitude are shown in Figure 7(b). Based on the analysis, it is observed that smaller vehicles (MTW) tend to have higher lateral amplitude at low flow conditions (Flow 1), with a mean of 8 m over the entire section. As traffic flow increases, the lateral amplitude decreases, and the distribution peak tends to shift. For Flow 2, the peak is about 6 m and in Flow 3 is about 3 m. The MTW tends to enjoy higher lateral freedom at free-flow conditions over the space, decreasing as traffic flow increases.

On the other hand, for the MThW-Cars class, at Flows 1 and 2, the lateral amplitude distribution was similar with a peak around 4 m. Considering the lane width of 3.5 m, this indicates that these vehicles tend to shift lanes, unlike MTW. Whereas at Flow 3 (near capacity), the lateral amplitudes dropped like MTW. Interestingly, in heavy vehicles (mainly Trucks and LCV) at all flow conditions, the lateral amplitude distributions tend to be similar at all flow conditions. This shows that heavy vehicles tend to have much less lateral movements over the road space with less lateral amplitude due to their sizes and limited lateral maneuverability.

Based on the lateral amplitude analysis, it is concluded that smaller vehicles tend to have higher lateral movement than heavy vehicles in the traffic stream. Further, due to the constrained movement, the lateral amplitude for the vehicles is reduced with the increase in flow levels.

Concluding remarks

Given the dataset constraints, driving behavior for mixed non-lane-based traffic conditions has not been much explored in comparison to homogeneous traffic conditions. This study was inspired by the importance of trajectory NGSIM datasets and the vast array of studies that it made possible. Extended trajectory data were developed in this study using a new stitching-based algorithm that can be considered a primary and novel attempt in developing a trajectory dataset for heterogeneous conditions. Through this research work, the following inferences were made:

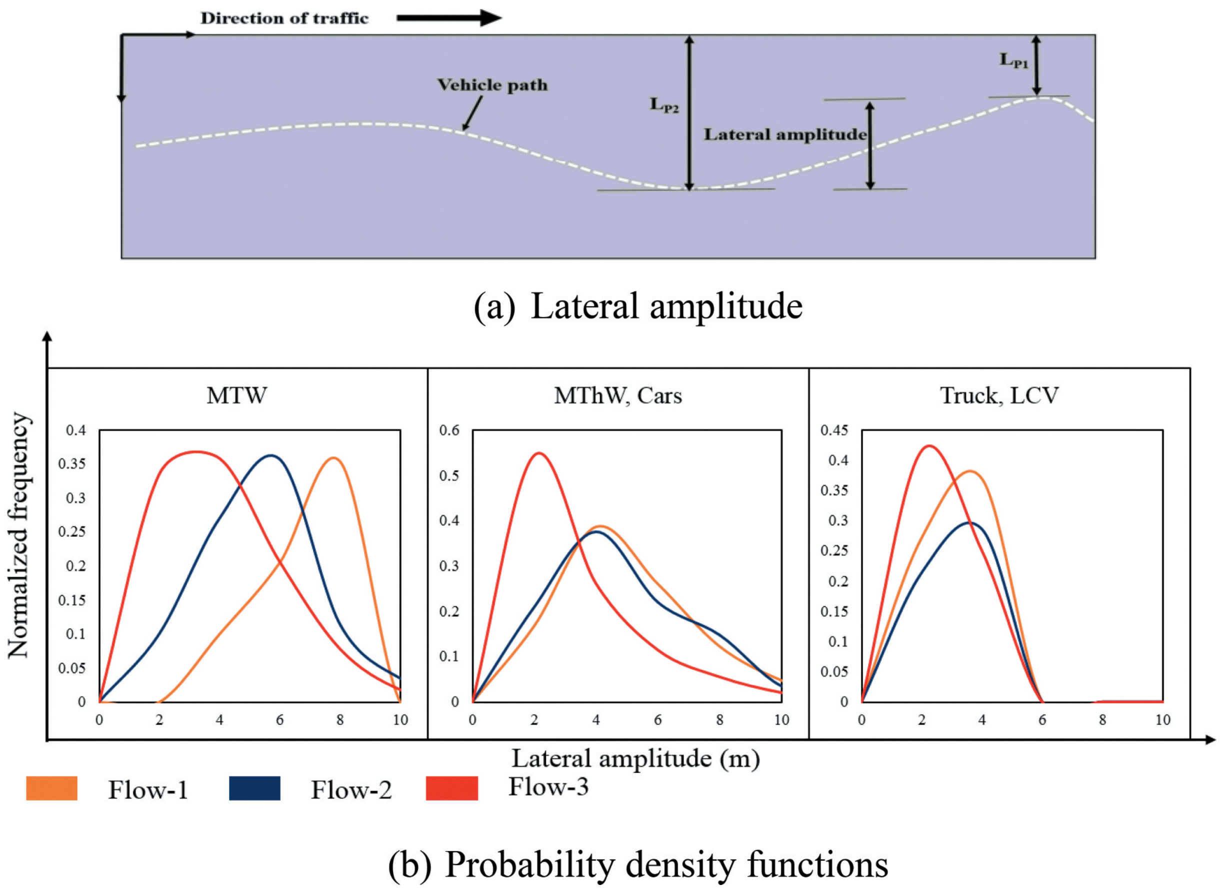


Figure 7. Lateral amplitude and its probability density function.

- Given the limitations of automation in the development of trajectory data with non-lane-based movements of vehicles and the variety of their sizes, the development of such data was a constraint in the past. Based on the proposed stitching algorithm, an extended trajectory dataset may no longer be a constraint in the future for conducting comprehensive studies on driving behavior. With the meticulousness of trajectory data, the accuracy of the developed stitching algorithm can be improved.
- The study also addressed the trajectory data development under blind spots. For example, a 10-m under the overpass acted as a blind spot for trajectory data development in the present case. In this case, using a high-degree polynomial equation, the missing points are predicted. Based on this logic, even the occlusion effect of vehicles, for example, smaller vehicles following a heavy vehicle, may be shadowed for short durations during the surveys. This trajectory data points for the smaller vehicles for that short duration may not be traced initially. Later, using this polynomial function, the missing points were predicted for those vehicles, and the accuracy was improved.
- With the limited availability of trajectory data from mixed traffic conditions, unlike NGSIM data for homogeneous traffic, the driving behavior under mixed traffic conditions has not been explored. Based on the developed stitching algorithm, trajectory data from mixed traffic conditions was quickly developed. Further, the developed trajectory datasets can help explore different aspects of driving behavior, including vehicle following, lateral movement, and traffic safety analytics, such as time to collisions and deceleration rate to avoid collisions.
- Based on the driving behavior investigation using lateral amplitude, it was found that smaller vehicles in the traffic stream tend to have greater lateral freedom than heavy vehicles. Further, with an increase in traffic-flow level, the lateral freedom has dropped along the road for those vehicles. Interestingly, for heavy vehicles, the lateral behavior was the same, irrespective of the traffic-flow level.
- Considering the dominance of lateral behavior in mixed traffic, this behavior has been investigated and revealed various driving phenomena in the traffic stream. The study demonstrates comprehensive efforts required to develop an extended trajectory dataset for mixed traffic conditions like NGSIM in the United States. As such, the proposed methodology and dataset should be of interest to researchers in developed and developing countries.
- In the present study, the thresholds for stitching were increased progressively for stitching the trajectory data. Still, in this direction, as a future research scope, more studies are required to standardize the stitching thresholds to limit the miss-stitches and improve the precision of the stitching process.

Study contributions

Given the advancements in technology over the last two decades, researchers identified the importance of vehicular trajectory data. Currently, traffic flow modeling concepts and traffic microsimulation highly rely on the trajectory data for the analysis. With the data constraints, the driving behavior from heterogeneous non-lane-based traffic conditions has not been explored much, unlike homogeneous traffic conditions. In the present work, extended trajectory data was developed by stitching logic and can be considered a primary and a novel attempt to develop trajectory data under heterogeneous traffic environments and be further beneficial for studying driving behavior to a greater extent. Besides this, the study developed exhaustive comprehensive extended trajectory data for Indian traffic conditions. Overall, the study can be considered as a unique contribution, keeping in view non-lane-based heterogeneous (weak-disciplined) and complex traffic flow interactions involving multiple classes of vehicles. In a way forward, the methodological contribution from this research work may develop a more profound vision related to developing longitudinal and lateral driving behavioral models, collision avoidance, high-end calibrated traffic simulation models. It can be further contemplated that the developed logic of building extended trajectories can be very much useful in offering insights from heterogeneous traffic context.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Narayana Raju http://orcid.org/0000-0002-3561-5676
Shriniwas Arkatkar http://orcid.org/0000-0002-1804-9465

References

Ali, Y., Z. Zheng, and M. M. Haque. 2018. “Connectivity’s Impact on Mandatory Lane-changing Behaviour: Evidences from a Driving Simulator Study.” *Transportation Research Part C: Emerging Technologies* 93: 292–309. doi:10.1016/j.trc.2018.06.008.

Bharadwaj, N., P. Kumar, S. Arkatkar, A. Maurya, and G. Joshi. 2016a. *Traffic Data Analysis Using Image Processing Technique on Delhi – Gurgaon Expressway. Current Science*110 (5): 808–822. doi:10.18520/cs/v110/i5/808-822.

Bharadwaj, N., S. Mathew, A. Pani, S. Arkatkar, G. Joshi, and K. Ravinder. 2016b. “Effect of Traffic Composition and Emergency Lane on Capacity: A Case Study of Intercity Expressway in India.” *Transportation Letters* 7867 (December): 1–17. doi:10.1080/19427867.2016.1265237.

Chen, D., J. Laval, Z. Zheng, and S. Ahn. 2012. “A Behavioral Car-following Model that Captures Traffic Oscillations.” *Transportation Research Part B: Methodological* 46 (6): 744–761. doi:10.1016/j.trb.2012.01.009.

Chiu, Y. C., L. Zhou, and H. Song. 2010. “Development and Calibration of the Anisotropic Mesoscopic Simulation Model for Uninterrupted Flow Facilities.” *Transportation Research Part B: Methodological* 44 (1): 152–174. doi:10.1016/j.trb.2009.06.001.

Chunchu, M., R. R. Kalaga, and N. V. S. K. Seethepalli. 2010. “Analysis of Microscopic Data under Heterogeneous Traffic Conditions.” *Transport* 25 (3): 262–268. doi:10.3846/transport.2010.32.

Drews, F. A., H. Yazdani, C. N. Godfrey, J. M. Cooper, and D. L. Strayer. 2009. “Text Messaging During Simulated Driving.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 51 (5): 762–770. doi:10.1177/0018720809353319.

Durrani, U., C. Lee, and H. Maoh. 2016. “Calibrating the Wiedemann’s Vehicle-following Model Using Mixed Vehicle-pair Interactions.” *Transportation Research Part C: Emerging Technologies* 67: 227–242. doi:10.1016/j.trc.2016.02.012.

Feng, W., A. Bigazzi, S. Kothuri, and R. Bertini. 2010. “Freeway Sensor Spacing and Probe Vehicle Penetration.” *Transportation Research Record: Journal of the Transportation Research Board* 2178 (1): 67–78. doi:10.3141/2178-08.

FHWA(Federal Highway Administration). 2007. *Next Generation Simulation (NGSIM) Overview*. The United States: weblink of FHWA.

Forbes, T. W., M. J. Zagorski, E. L. Holshouser, and W. A. Deterline. 1959. “Measurement of Driver Reaction to Tunnel Conditions.” *Proceedings of the Highway Research Board* 35: 345–357.

Gazis, D. C., R. Herman, and R. B. Potts. 1959. “Car-Following Theory of Steady-State Traffic Flow.” *Operations Research* 7 (4): 499–505. doi:10.1287/opre.7.4.499.

Gipps, P. G. 1981. “A Behavioural Car-following Model for Computer Simulation.” *Transportation Research Part B* 15 (2): 105–111. doi:10.1016/0191-2615(81)90037-0.

Hao, H., W. Ma, and H. Xu. 2016. “A Fuzzy Logic-based Multi-agent Car-following Model.” *Transportation Research Part C: Emerging Technologies* 69: 477–496. doi:10.1016/j.trc.2015.09.014.

Haque, M. M., and S. Washington. 2015. “The Impact of Mobile Phone Distraction on the Braking Behaviour of Young Drivers: A Hazard-based Duration Model.” *Transportation Research Part C: Emerging Technologies* 50: 13–27. doi:10.1016/j.trc.2014.07.011.

Hidas, P. 2005. “Modelling Vehicle Interactions in Microscopic Simulation of Merging and Weaving.” *Transportation Research Part C: Emerging Technologies* 13 (1): 37–62. doi:10.1016/j.trc.2004.12.003.

Huang, Y., R. Bird, and M. Bell. 2009. “A Comparative Study of the Emissions by Road Maintenance Works and the Disrupted Traffic Using Life Cycle Assessment and Micro-simulation.” *Transportation Research Part D: Transport and Environment* 14 (3): 197–204. doi:10.1016/j.trd.2008.12.003.

Jenelius, E., and H. N. Koutsopoulos. 2013. “Travel Time Estimation for Urban Road Networks Using Low Frequency Probe Vehicle Data.” *Transportation Research Part B: Methodological* 53: 64–81. doi:10.1016/j.trb.2013.03.008.

Johansson, G., and K. Rumar. 1971. “Drivers’ Brake Reaction Times.” *Human Factors: The Journal of Human Factors and Ergonomics Society* 13 (1): 23–27. doi:10.1177/001872087101300104.

Kanagaraj, V., G. Asaithambi, T. Toledo, and T.-C. Lee. 2015. “Trajectory Data and Flow Characteristics of Mixed Traffic.” *Transportation Research Record: Journal of the Transportation Research Board* 2491 (1): 1–11. doi:10.3141/2491-01.

Kesting, A., M. Treiber, M. Schönhof, and D. Helbing. 2008. “Adaptive Cruise Control Design for Active Congestion Avoidance.” *Transportation Research Part C: Emerging Technologies* 16 (6): 668–683. doi:10.1016/j.trc.2007.12.004.

Kilpeläinen, M., and H. Summala. 2007. “Effects of Weather and Weather Forecasts on Driver Behaviour.” *Transportation Research. Part F, Traffic Psychology and Behaviour* 10 (4): 288–299. doi:10.1016/j.trf.2006.11.002.

Krishnamoorthy, K. 2016. “Wilcoxon Rank-Sum Test.” In *Handbook of Statistical Distributions with Applications*. 376–378. New york, The united states: CRC Press, Taylor and Francis group. doi:10.1201/9781420011371-35.

Kumar, P., M. S. Bains, N. Bharadwaj, S. Arkatkar, and G. Joshi. 2020. “Impact Assessment of Driver Speed Limit Compliance Behavior on Macroscopic Traffic Characteristics under Heterogeneous Traffic Environment.” *Transportation Letters* 12 (1): 54–65. doi:10.1080/19427867.2018.1512200.

Leclercq, L., N. Chiabaut, J. Laval, and C. Buisson. 2007. “Relaxation Phenomenon after Lane Changing: Experimental Validation with NGSIM Data Set.” *Transportation Research Record: Journal of the Transportation Research Board* 1999 (1): 79–85. doi:10.3141/1999-09.

Lee, G., S. Joo, C. Oh, and K. Choi. 2013. “An Evaluation Framework for Traffic Calming Measures in Residential Areas.” *Transportation Research Part D: Transport and Environment* 25: 68–76. doi:10.1016/j.trd.2013.08.002.

Lu, X., and A. Skabardonis. 2007. “Freeway Traffic Shockwave Analysis: Exploring NGSIM Trajectory Data.” In *Transportation Research Board 86th Annual Meeting, January 2007*, 19.Washington DC, The United States: Transportation research board.

Montanino, M., and V. Punzo. 2013. “Making NGSIM Data Usable for Studies on Traffic Flow Theory.” *Transportation Research Record: Journal of the Transportation Research Board* 2390 (1): 99–111. doi:10.3141/2390-11.

Munigety, C., V. Vicraman, and T. Mathew. 2014. “Semiautomated Tool for Extraction of Microlevel Traffic Data from Videographic Survey.” *Transportation Research Record: Journal of the Transportation Research Board* 2443 (1): 88–95. doi:10.3141/2443-10.

Newell, G. F. 2002. “A Simplified Car-following Theory: A Lower Order Model.” *Transportation Research Part B: Methodological* 36 (3): 195–205. doi:10.1016/S0191-2615(00)00044-8.

Papailias, F., and D. D. Thomakos. 2015. “An Improved Moving Average Technical Trading Rule.” *Physica A: Statistical Mechanics and Its Applications* 428: 458–469. doi:10.1016/j.physa.2015.01.088.

Pipes, L. A. 1953. “An Operational Analysis of Traffic Dynamics.” *Journal of Applied Physics* 24 (3): 274–281. doi:10.1063/1.1721265.

Raju, N., P. Kumar, A. Jain, S. S. Arkatkar, and G. Joshi. 2018. "Application of Trajectory Data for Investigating Vehicle Behavior in Mixed Traffic Environment." *Transportation Research Record* 97. doi:10.1177/0361198118787364.

Raju, N., P. Kumar, C. Reddy, S. Arkatkar, and G. Joshi. 2017. "Examining Smoothening Techniques for Developing Vehicular Trajectory Data under Heterogeneous Conditions." *Journal of the Eastern Asia Society for Transportation Studies* 12: 1549–1568.

Raju, N., S. Arkatkar, and G. Joshi. 2020. "Effect of Construction Work Zone on Traffic Stream Parameters Using Vehicular Trajectory Data under Mixed Traffic Conditions." *Journal of Transportation Engineering, Part A: Systems* 146 (6): 5020002. doi:10.1061/JTEPBS.0000353.

Remias, S., A. Hainen, C. Day, T. Brennan, H. Li, E. Rivera-Hernandez, J. Sturdevant, S. Young, and D. Bullock. 2013. "Performance Characterization of Arterial Traffic Flow with Probe Vehicle Data." *Transportation Research Record: Journal of the Transportation Research Board* 2380 (1): 10–21. doi:10.3141/2380-02.

Shirke, C., N. Sumanth, S. Arkatkar, A. Bhaskar, and G. Joshi. 2017. "Modeling Expressway Lane Utilization and Lane Choice Behaviour: A Case Study on Delhi–Gurgaon Expressway." *Transportation Letters*. doi:10.1080/19427867.2017.1336859.

Sobhani, A., W. Young, and M. Sarvi. 2013. "A Simulation Based Approach to Assess the Safety Performance of Road Locations." *Transportation Research Part C: Emerging Technologies* 32: 144–158. doi:10.1016/j.trc.2012.10.001.

Toledo, T., O. Cats, W. Burghout, and H. N. Koutsopoulos. 2010. "Mesoscopic Simulation for Transit Operations." *Transportation Research Part C: Emerging Technologies* 18 (6): 896–908. doi:10.1016/j.trc.2010.02.008.

Underwood, G., D. Crundall, and P. Chapman. 2011. "Driving Simulator Validation with Hazard Perception." *Transportation Research. Part F, Traffic Psychology and Behaviour* 14 (6): 435–446. doi:10.1016/j.trf.2011.04.008.

Vicraman, V., C. Ronald, T. Mathew, and K. V. Rao (2014). "Traffic Data Extractor." IIT, Bombai. http://www.civil.iitb.ac.in/tvm/tde2

Wiedemann, R. 1974. *Simulation des Straßenverkehrsflusses. Schriftenreihe des IfV, 8, Institut für Verkehrswesen*. Universität Karlsruhe, Germany: Universität Karlsruhe.