

ORIGINAL PAPER

Open Access



# Flexible car-following models for mixed traffic and weak lane-discipline conditions

Vasileia Papathanasopoulou<sup>1\*</sup>  and Constantinos Antoniou<sup>2</sup>

## Abstract

Heterogeneous mixture of vehicle types and lack of lane discipline are common characteristics of cities in the developing countries. These conditions lead to driving manoeuvres that combine both longitudinal and lateral movements. Modeling this driving behavior tends to be complex and cumbersome, as various phenomena, such as multiple-leader following, should be addressed. This research attempts to simplify mixed traffic modeling by developing a methodology, which is based on data-driven models. The methodology is applied on mixed traffic, weak lane-discipline trajectory data, which have been collected in India. A well-known car-following model, Gipps' model, is also applied on the same data and is used as a reference benchmark. Regarding the lateral manoeuvres, the focus is given on identification of significant lateral changes, which could indicate a lane-changing situation. Methods that allow monitoring structural changes in regression models could be used for this purpose. The ability of capturing lane changes is explored. A typical example is illustrated and further discussion is motivated.

**Keywords:** Mixed traffic, Weak lane discipline, Data-driven models, Machine-learning, Virtual lanes

## Introduction

Traffic simulation models have been typically formulated for lane-based conditions and European traffic. However, simulation of mixed traffic flow in weak lane based heterogeneous conditions poses additional challenges. In recent years, there has been an increasing interest in modeling driving behavior in developing countries where conditions, such as non-lane discipline and heterogeneity in vehicle types, prevail. Wong et al. [1] have explored traffic characteristics of mixed traffic flows in urban arterials and have focused on motorcycles, the proportion of which is high in Asian countries. Traffic flow in the developing countries is very complex in nature and safety issues arise.

Due to the lack of lane discipline, it is difficult to identify leader-follower pairs and to decide if a car-following or a lane changing model should be applied. This research aims to provide some more input into this ongoing active research field. Car-following and lane-changing models describe the longitudinal and lateral movements of drivers. However, these two behavioral models may not be

able to describe the integrated driving behavior independently [2]. Lateral interactions take place along with longitudinal interactions in mixed traffic conditions. There have been many attempts to model these behaviors separately. Some of these attempts found in the literature are described in the next section.

This research was motivated by several considerations. Car-following models, which replicate the behavior of a driver following another vehicle, are widely used in the deployment of traffic simulation models. However, only fewer studies have focused on mixed traffic conditions. Due to the complex driver behavior and vehicular interactions and manoeuvres it is difficult to model the traffic flow through analytical methods [3]. Modeling driving behavior in mixed traffic streams is still a challenge. Heterogeneous mixture of vehicle types and violation of lane regulations are common characteristics in cities in developing countries. These characteristics are difficult to be simulated using conventional microscopic models. In cases of car-following situations, there is difficulty in the determination of leader-follower pairs due to multiple-leader following. Furthermore, in cases of lane changing situations there is difficulty in the determination of lanes, as drivers do not obey the real lane marks. To overcome some of the associated limitations, in this research a

\*Correspondence: [vasileia.papathanasopoulou@gmail.com](mailto:vasileia.papathanasopoulou@gmail.com)

<sup>1</sup>School of Rural and Surveying Engineering, National Technical University of Athens, 9, Iroon Polytechniou St., Zografou Campus, 15780 Athens, Greece  
Full list of author information is available at the end of the article

methodology is proposed using temporary virtual lanes in order to capture heterogeneity in vehicle width and speed.

The existing approaches for mixed traffic conditions do not adapt dynamically to the current conditions. Lanes, strips or cells with a predefined width [2, 4], which are used to simulate mixed traffic, do not ensure that the appropriate width has been selected, as half vehicle or two vehicles may fit into this width. Heterogeneity in vehicle types lead to various widths of virtual lanes and various speeds. On the other hand, temporary virtual lanes allow only one vehicle to fit in each lane.

The main objectives of this research are:

- to explore the feasibility of modeling mixed traffic conditions using data-driven models
- to compare the performance of data-driven models versus conventional models, in particular Gipps model
- to estimate model efficiency considering the difference in following behavior across different vehicle pairs
- to introduce the concept of temporary virtual lanes based on identification of significant lateral changes.

An integrated methodology is developed for modeling mixed traffic conditions. The focus is given on data-driven car-following models and the identification of significant lateral positions that may be indicative of the traffic situation of a vehicle (car-following, lane-changing or free flow). In the case study, a data-driven model has been used for speed estimation using mixed traffic trajectory data from India. Then, lateral manoeuvres are investigated using an algorithm for identification of structural changes in data. Finally, issues for further analysis and future prospects are discussed.

## Literature review

Asaithambi et al. [5] have reviewed driver behavior models under mixed traffic conditions and have pointed out limitations of current models, arguing that the main limitation is that they do not explicitly consider the wider range of situations that drivers in mixed traffic face. Munigety and Mathew [2] have identified that due to weak lane discipline, drivers maneuvering in mixed traffic streams exhibit some peculiar patterns such as maintaining shorter headways, swerving, and filtering. They have also proposed that the lane should be divided into small strips in order to handle virtual lane movements. Li et al. [6] have proposed a car-following model that considers the effect of two-sided lateral gaps and they have shown that their model has larger stable region compared to a car-following model that captures the impacts from the lateral gap on only one side. In addition, Parsuvanathan [7] used proxy lanes between the main lanes.

It is assumed that free space is perceived as lanes by small vehicles. However, distribution and types of vehicles could affect the width of the lanes. A grid-based modeling approach akin to cellular automata [8] and a strip-based modelling method [4] have also been proposed. Mathew et al. [4] have based their idea on portions of traffic queues instead of regular main lane queues. Kanagaraj et al. [3] have evaluated the performance of different car following models under mixed traffic conditions. However, they have not taken into account the fact that a vehicle may not be exactly in line with its leading vehicle due to weak lane discipline in mixed traffic. Metkari et al. [9] have modified an existing car-following model in order to take into account lateral movements and include mixed traffic conditions. Choudhury and Islam [10] have developed a latent leader acceleration model.

Maurya [11] developed comprehensive driver behavior model which considers concurrently both longitudinal and lateral interaction with roadway and traffic features. Chunchu et al. [12] analyzed vehicle composition, lateral distribution of vehicles, lateral gaps and longitudinal gaps, in mixed traffic stream. Lan and Chang [13] used General Motors model to simulate the motorcycle's following behaviors in two cases: (1) only one leading vehicle in front; (2) two or more leading vehicles in front and neighboring-front (including left-front, right-front or both). The present research attempts to cover more cases. The problem of dealing with non-lane discipline conditions has been treated either by splitting lanes into small strips [4] or small cells using cellular automata model [14]. Furthermore, a veering angle and a path selection have been used to update the lateral position [15, 16]. Social Force models and friction forces [17] have also been proposed. Considering the existing literature review, the concept of temporary virtual lanes is innovative and flexible enough to adapt to mixed traffic conditions.

## Methodology

In this section, a methodology is developed for simulation of mixed traffic conditions. Mixed traffic flow is considered when speed differential among different types of vehicles is quite substantial and the desired number of overtaking increases with limited opportunities to overtake [18]. Mehar et al. [19] define as mixed traffic conditions when there are several categories of vehicles sharing and moving on same carriageway width without any physical segregation between motorized and non-motorized vehicles, and without proper lane discipline. Due to the wide variations in physical dimensions and speeds of various vehicles, it is difficult to impose lane discipline. The vehicles occupy any available lateral position on the road space, while the small vehicles, such as motor cycles often utilize gaps between larger vehicles in the traffic stream.

In mixed traffic flow, there are different combinations of vehicles for leader–follower pairs [3].

This study focuses on heterogeneity of vehicle width by considering temporary virtual lanes with different width. On the other hand, heterogeneity of vehicle speed is also critical and is considered by using data–driven models that use as input speeds. For example, three wheeler passenger vehicles and goods vehicles are having almost similar width in shape, but their speeds and driver behavior could be completely different.

**Virtual lanes and leader–follower pair identification**

**Determination of virtual lanes**

A typical example of modification of virtual lane change is illustrated in Fig. 1. In this figure, there are two vehicles. The first vehicle follows the virtual lane  $i$ . While there are small lateral movements, it is considered that it does not change lane. However, when its movement is constrained by the hatched vehicle at the breakpoint, it is considered that it changes lane and then follows virtual lane  $i + 1$ . The challenge is that vehicles are moving constantly laterally. This could be addressed in two distinct ways. The first one is to estimate the threshold that indicates a lane change. The second one is using change detection algorithms. In this research the focus is given on the second approach, namely on identifying significant changes in lateral positions, so as the appropriate microscopic model to be applied. Algorithms that are capable of finding major changes in data sequence could be used.

Heterogeneity in vehicle types implies various widths of vehicles and thus various widths of virtual lanes. The width of a temporary virtual lane  $W$  could be estimated by Eq. 1, if no significant lateral changes and breakpoints are identified. The estimation of temporary virtual lanes is also illustrated in Fig. 2.

$$W = \max(x_t, x_{t+1} + \dots + x_{t+n}) - \min(x_t, x_{t+1} + \dots + x_{t+n}) + w_v \tag{1}$$

where  $x_t$  is the position of the center of the vehicle, measured from the left–most side of the roadway for each time instant  $t+i$  and  $w_v$  is the width of the vehicle.

The estimation procedure of virtual lane width takes place between two consecutive breakpoints. The same procedure could be applied in all types of urban carriageways. However, the sensitivity of the algorithm, which identifies major changes in data sequence, should be set to adapt conditions of the respective road network. For instance, on a highway, larger lateral movements are expected to imply a lane change manoeuvre.

**Identification of leader–follower vehicles**

The probability of a given front vehicle to be the governing leader depends on the type of the lead vehicle and the extent of lateral overlap with the following vehicle [10].

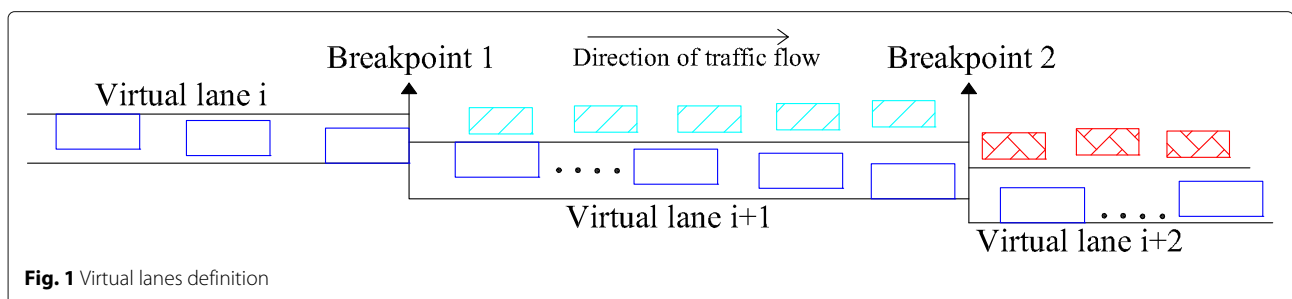
In order to apply a microscopic model, it should be determined whether there is a vehicle pair of follower–leader. The main characteristic of mixed traffic is that the size of overlap between the leader and the follower varies. Assuming that the lateral and longitudinal coordinates of the front center of each vehicle ( $x_{c_i}, x_{c_i}$ ) are known, it could be defined which vehicle follows the other. The coordinates for the left and the right lateral bound of each vehicle are estimated per time instant  $t$  by Eqs. 2 and 3 (as shown in Fig. 3a).

$$x_{l_i}(t) = x_{c_i}(t) - \frac{w_i}{2} - s_i(t) \tag{2}$$

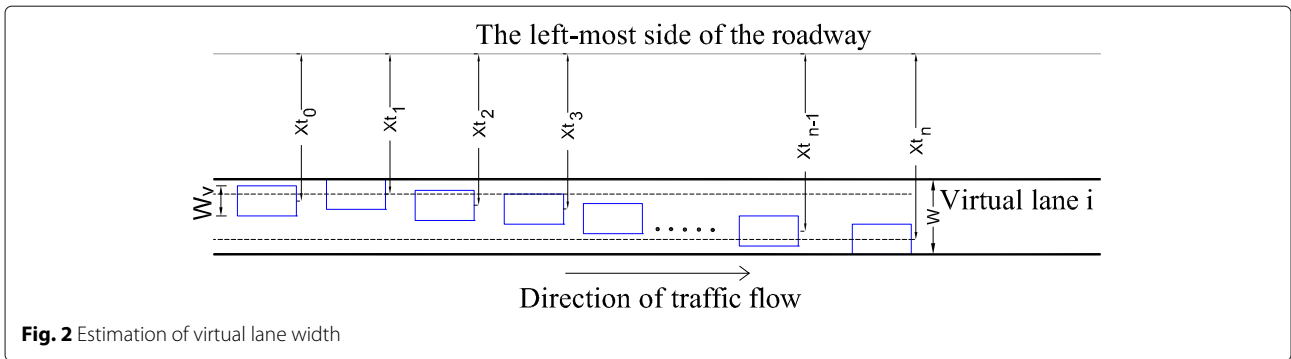
$$x_{r_i}(t) = x_{c_i}(t) + \frac{w_i}{2} + s_i(t) \tag{3}$$

where  $i: 0,1,2,n$  vehicle index  $x_{c_i}$ : lateral coordinate of the front center of vehicle  $i$ ,  $x_{l_i}$ : lateral coordinate of the front left bound of vehicle  $i$ ,  $x_{r_i}$ : lateral coordinate of the front right bound of vehicle  $i$ ,  $w_i$ : width of vehicle  $i$   $s_i$ : a lateral safety distance for vehicle  $i$ .

In order to define the car–following vehicle pairs, the longitudinal position of the leader should be in front of the following vehicle and in a distance  $L$  that could influence the movement of the following vehicle (Eq. 4). In addition, a part of the front side of a vehicle should overlap a part of the front side of another vehicle (Eq. 5). This overlap is evident in Fig. 3b with light blue color. Each vehicle  $i$  is considered as follower and then a leader vehicle is required



**Fig. 1** Virtual lanes definition



to fulfill the conditions, described by Eqs. 4 and 5, at the same instant  $t$ :

$$y_{follower}(t) \leq y_{leader}(t) \leq y_{follower}(t) + L \quad (4)$$

$$\begin{aligned} x_{l_{follower}}(t) &\leq x_{r_{leader}}(t) \\ x_{l_{leader}}(t) &\leq x_{r_{follower}}(t) \end{aligned} \quad (5)$$

A scenario with two leaders and one follower case is also possible. For instance, a bus could be the follower and a part of its front side may overlap with two leaders such as two motorcycles or a small vehicle and a motorcycle. In this case the closest vehicle according to the direction of movement is chosen as the most critical leader [20]. If no vehicles are identified as leaders, then the driving situation of the vehicle is free flow.

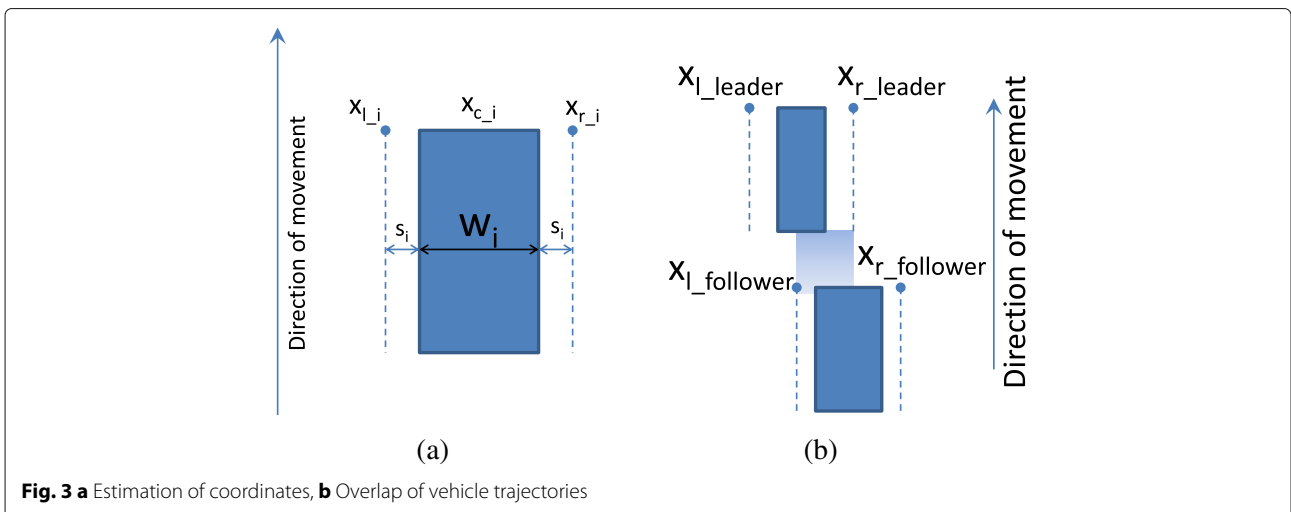
**Operationalization process**

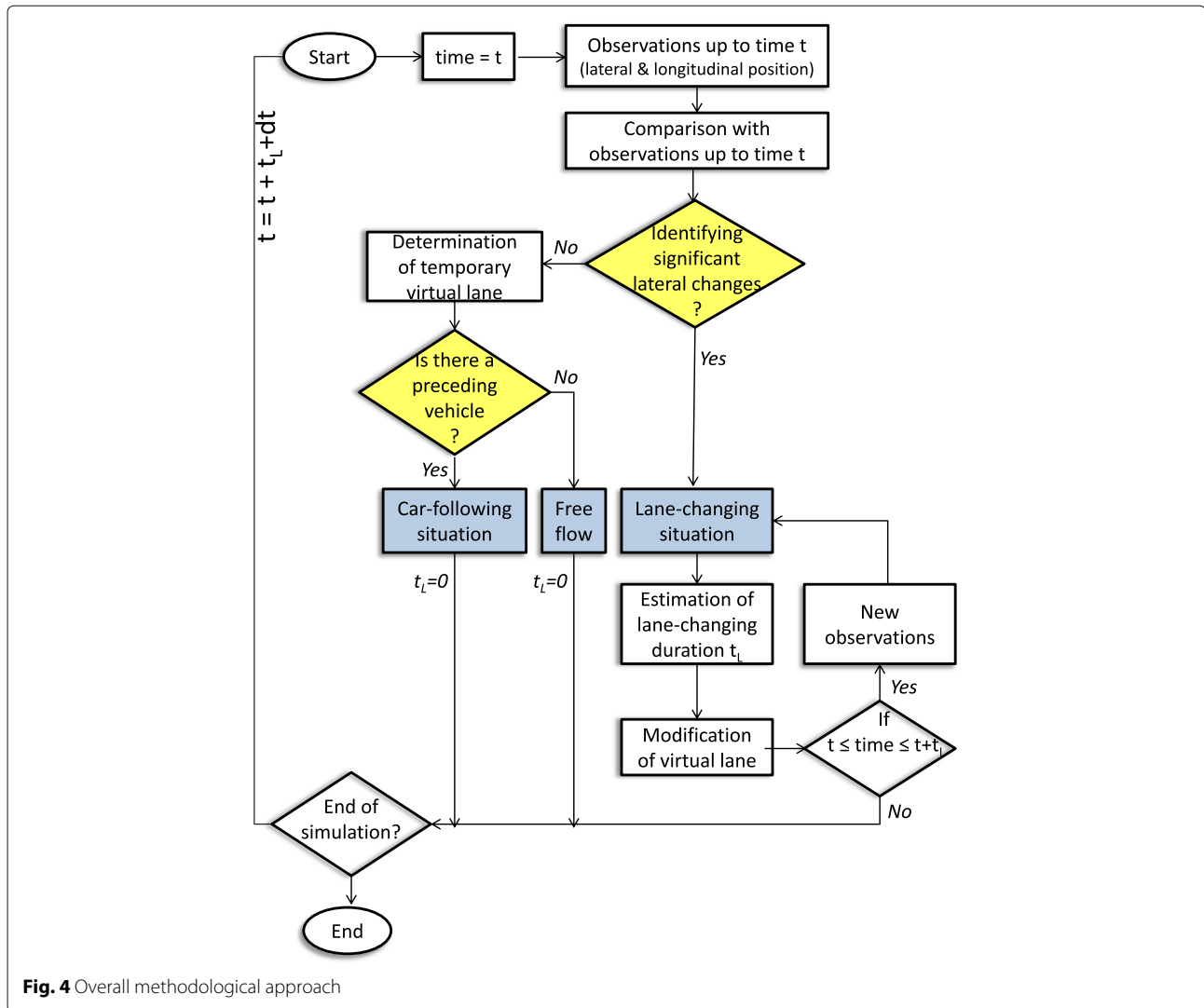
It is assumed that all vehicles are moving without lane discipline. In order to simplify this traffic situation, temporary virtual lanes for each vehicle are defined. The methodology is based on the idea that each driver follows his own temporary virtual traffic lane until his lane over-

laps with the virtual lane of another driver and thus he is forced to modify it. The proposed methodological approach is outlined in Fig. 4. Longitudinal and lateral positions are recorded per time instant and saved in a database. Then significant lateral changes are identified using appropriate algorithms that allow monitoring structural changes in linear regression models. If no significant lateral change is identified then lateral information is used for determination of a temporary virtual lane and then a car-following model or a free flow model is applied if at least one preceding vehicle is identified or not respectively. For identification of the front vehicle more details are provided in the next subsection. On the other hand, if a breakpoint is observed in data sequence, namely if significant lateral changes are identified, then a lane-changing situation is indicated and the virtual lane needs to be modified. A lane-changing model should be applied for time  $t_L$ , time of lane-changing duration. Then the process is iterated for the following time instants.

**Data-driven modeling**

The process for data-driven model development is outlined in Fig. 5. The approach includes two parts: training and application. First the required explanatory variables





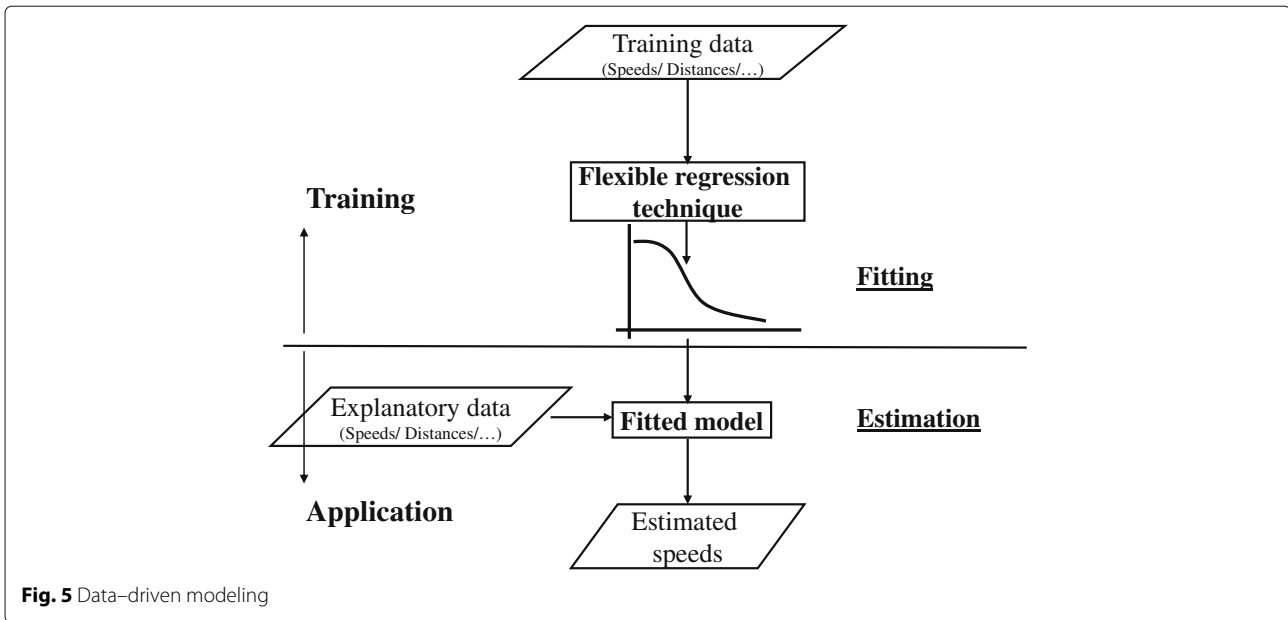
of the model are determined and the appropriate surveillance data are collected. In the training step traffic models are estimated according to the available surveillance data using a flexible regression technique, while in the application step the fitted model is applied to provide predictions using new observations.

Estimation has been achieved without assuming any predefined functional form; instead a flexible regression method. Various machine learning techniques could be used in this context. Other data-driven methods, including neural networks [21], Gaussian processes [22] and Kernel methods offering similar capabilities, have also been used in applications [23]. In this research locally weighted regression has been used, as it comprises much of the simplicity of linear least squares regression with the flexibility of nonlinear regression.

Locally weighted regression (loess) could be considered as a generalization of the k-nearest neighbor method

[24]. It was firstly introduced by Cleveland [25] and the following analysis is based on [26].

Locally weighted regression  $y_i = g(x_i) + \epsilon_i$ , where  $i = 1, \dots, n$  index of observations,  $g$  is the regression function and  $\epsilon_i$  are residual errors, provides an estimate  $g(x)$  of each regression surface at any value  $x$  in the  $d$ -dimensional space of the independent variables. Correlations between observations of the response variable  $y_i$  and the vector with the observations  $d$ -tuples  $x_i$  of  $d$  predictor variables are identified. Local regression provides an estimation of function  $g(x)$  near  $x = x_0$  according to its value in a particular parametric class. This estimation could be achieved by adapting a regression surface to the data points within a neighborhood of the point  $x_0$ , which is bounded by a smoothing parameter: span. The span determines the percentage of data that are considered for each local fit and hence the smoothness of the estimated surface is influenced [27]. The span ranges from 0 (wavy curve) to 1



(smooth curve). Each local regression uses either a first or a second degree polynomial that is specified by the value of the “degree” parameter of the method (degree=1 or degree=2).

The data are weighted according to their distance from the center of neighborhood  $x$ , therefore a distance and a weight function are required. As a distance function  $p$ , Euclidean distance could be used for a single independent variable; otherwise, for the multiple regression case, any variable should be evaluated on a scale before applying a standard distance function [28]. A weight function defines the size of influence on fit for each data point taking for granted that nearby points have higher influence than the most distant. Therefore the weight function calculates the distances between each point and the estimation point and higher values in a scale from 0 to 1 are set for the nearest observations. A weight function should meet the requirements determined by Cleveland [25] and the most common one is the tri-cube function:

$$W(u) = \begin{cases} (1 - u^3)^3, & 0 \leq u \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

The weight of each observation  $(y_i, x_i)$  is defined as following:

$$w_i(x) = W[p(x, x_i)/d(x)] = \left(1 - \left(\frac{x_i - x}{d(x)}\right)^3\right)^3 \quad (7)$$

where  $d(x)$  is the distance of the most distant predictor value within the area of influence. In the loess method, weighted least squares are used so as linear or quadratic functions of the independent variables could be fitted at

the centers of neighborhoods [25]. The objective function that should be minimized is:

$$\sum_{n=1}^n w_i \cdot \epsilon_i^2 \quad (8)$$

### Evaluation

The performance of the models presented in this paper is evaluated using the normalized root mean square error RMSN [29]. The RMSN assesses the overall error and performance of each method estimating the difference between the observed values  $Y_n^{obs}$  and their simulated counterparts  $Y_n^{sim}$ . It is calculated from the following equation:

$$RMSN = \frac{\sqrt{N \cdot \sum_{n=1}^N (Y_n^{sim} - Y_n^{obs})^2}}{\sum_{n=1}^N Y_n^{obs}} \quad (9)$$

### Case study set-up

#### Data collection

In order to evaluate the feasibility of the methodological framework on mixed traffic trajectory data, data collected in India were used [30]. The video data were collected on a six-lane separated urban arterial road at the Maraimalai Adigalar Bridge in Saidapet, Chennai, India. The section was on a bridge, which ensured that the road geometry was uniform and that there were no nearby intersections, bus stops, parked vehicles, or other side factors that could affect drivers’ behavior. Furthermore, there was no interaction between the vehicle traffic and pedestrians, because the pedestrian walkway is segregated by a barrier. A detailed description of the data could be found in [30]. The data are presented in two parts- two excel files

for the data collected in the periods 2:45–3:00 PM and 3:00–3:15 PM, on February 13, 2014. Each excel sheet contains columns of variables, such as time, vehicle type, length and width, longitudinal position, speed, acceleration and lateral position, speed, acceleration. Longitudinal position is the position of the front of the vehicle, measured from the upstream end of the section, while lateral position is the position of the center of the vehicle, measured from the left-most side of the roadway. The trajectory data are available at the address <http://toledo.net.technion.ac.il/downloads/>.

### Data processing

First, data were organized in ascending order of vehicle ID, so as the trajectory of each vehicle to be continuous and observations of other vehicles not to interfere. Then, only observations appropriate for microscopic analysis are selected (flag=0). As coordinates of the front center of each, longitudinal and lateral positions are used. Regarding the considered speed for each vehicle, the resultant speed is estimated by Eq. 10.

$$v_i(t) = \sqrt{v_{long_i}^2 + v_{lat_i}^2} \quad (10)$$

where  $v_i$ : resultant speed of vehicle  $i$ ,  $v_{long_i}$ : longitudinal speed of vehicle  $i$  and  $v_{lat_i}$ : lateral speed of vehicle  $i$ .

In addition, a new column is added which includes the observed speed for the next time instant, namely the speed that should be predicted for each observation. Actually this is the speed that corresponds to time  $t + 0.5$  s and to the same vehicle ID. If there is no observation for this vehicle and for the next time instant, NA is given. Afterwards, rows with NA in this column are omitted, as there is no observed speed to compare with the estimated one by the proposed methodology.

Due to the nature of mixed traffic data, the next step was to define the car-following sequence, namely which vehicle is in front of the other. [30] have identified that in 45% of the observations the overlap between the leader and the follower is less than half the follower width. The methodology described in section was adopted for the identification of the front vehicle. Observations that correspond to vehicles with no leading vehicle were excluded. As lateral safety distance,  $s = 0.20$  m is considered for each vehicle on both sides. As distance  $L$  in Eq. 4,  $L = 200$  m is considered. If no vehicles are identified as leaders, then these observations are omitted, as they do not correspond to car-following state. The same procedure was also used with the validation on dataset data300. Finally, dataset “data245” includes 47036 observations corresponding to 1511 vehicle pairs and dataset “data300” 45982 observations corresponding to 1488 vehicle pairs.

### Estimation of conventional models

There are several traffic micro-simulation packages, such as AIMSUN, PARAMICS, TransModeler and VISSIM, that could be used as a reference benchmark in terms of conventional models. AIMSUN utilizes a safety distance car-following model, the Gipps model, while PARAMICS uses the Fritzsche car-following model [31] and VISSIM is based on a psychophysical model. Mehar et al. [19] found that the VISSIM in its original form is not able to simulate mixed traffic conditions that prevail on Indian highways and proposed a method for model calibration appropriate for mixed traffic. A few modifications to the default behavioral parameters of VISSIM are required to effectively simulate Indian mixed traffic conditions [32]. Several studies have also demonstrated the use of VISSIM in simulating mixed traffic in different countries [33, 34]. On the other hand, VISSIM model contains the largest number of parameters which are also not easily interpreted to familiar driving factors such as the desired speed. The Fritzsche model of Paramics is similar to VISSIM model and includes the same number of parameters. However, AIMSUN is the model with the smallest number of parameters and the most interpretable ones, allowing the best possible results with less calibration work [35]. In addition, Kanagaraj et al. [3] evaluated four different car following models, in particular Gipps Model, Intelligent Driver Model (IDM), Krauss Model and Das and Asundi Model, under mixed traffic conditions and have shown that Gipps model is able to replicate the field conditions better than other models in non-steady state. Among the aforementioned models, Gipps’ model [36], which is used in AIMSUN, is considered for this case study. More traffic simulation models should be also tested as future prospect.

The Gipps model is used as reference in order to monitor and evaluate the effectiveness of the proposed method. This model requires as input the same data as the proposed method and thus a direct comparison would be feasible. First, a calibration of model parameters is required. There are six parameters in this model that have to be calibrated. The apparent reaction time is considered as 0.5 s and for calibration of the rest of parameters an optimization process is implemented. Dataset “data245” was used for calibration and “data300” for validation. The calibration process was performed within the R software for statistical computing [37]. In particular, the Improved Stochastic Ranking Evolution Strategy (ISRES) algorithm was used, which is included in the package “nloptr” [38] and is appropriate for nonlinearly constrained global optimization. This method is implemented in a simple way and supports arbitrary nonlinear inequality and equality constraints in addition to the bound constraints. In addition, it incorporates heuristics to escape local optima. The objective function that was minimized

is the RMSN between the observed and simulated values of speeds:

$$RMSN \left( v_{follower}^{obs}, v_{follower}^{sim} \right) \tag{11}$$

Bounds and initial values for model parameters have been defined in a previous work [39] and are shown in Table 2. These initial values have been defined as optimal values for data with lane discipline by algorithm ISRES in that research. Thus, it is expected that there will be a differentiation in optimal values due to different nature of data. Three samples of 5000 observations were selected randomly from dataset “data245”. The amount of observations used in each sample are summarized in Table 1 per vehicle type. A representative amount for each vehicle type is included in each sample. The optimization process was implemented for each sample separately and the results are presented in Table 2. For these samples the optimization process has converged to the optimal set of parameters after approximately 10,000 iterations. Using novel stochastic simulation and optimization approaches (such as using quasi-random Sobol sequences [40], instead of pseudo-random numbers) can reduce the required number of iterations and thus the overall computational burden [41].

For all samples similar parameter values have been produced and thus the optimization process for the whole dataset is considered unnecessary. Instead, the mean of the three optimized sets of parameters is selected and is presented in the last column of Table 2. Furthermore, the authors explored the impact of different initial values and the algorithm converged to the same solution, suggesting robustness of the optimization process. Looking into initial values that were appropriate for traffic under normal conditions and values optimized for mixed traffic conditions, the main difference is observed in maximum braking  $b$  that the driver of vehicle wishes to apply in order to avoid a crash. This could be attributed to the fact that more abrupt driving is observed in a mixed traffic environment. The minimum value of the objective function, namely the RMSN that was achieved with these optimal

values of parameters was 21%. Then, the calibrated model is validated on dataset “data300” and RMSN is estimated between observed and predicted speed per time instant. The results are shown in Fig. 6 and a comparison with the proposed method is feasible.

**Application of data-driven models**

In this research the explanatory variables per each time instant  $t$  have been considered as independent predictor variables for the estimation of the response variable (for instance speed) for the next time instant  $(t+\tau)$ , where  $\tau$  is the apparent reaction time. Estimation is achieved without assuming any predefined functional form; instead a flexible regression method can be used. The next step is the fitting of the proposed methodology for car-following situations using data-driven models. The problem to be addressed is the speed estimation of each vehicle, when the available data include its speed, the speed of the preceding vehicle and the distance between the two vehicles (in the previous time instant). Locally weighted regression could be used for the application. In the training step the flexible car-following model is fitted or calibrated on the surveillance data and validated on the other dataset.

**Exploration of data-driven car-following models**

The proposed method identifies the relationships between predictor variables  $v_{leader}(t)$ ,  $v_{follower}(t)$ , the distance  $D(t)$  between the two vehicles and the response data  $v_{follower}(t+\tau)$ , where  $\tau=0.5$  s. After the relevant pattern from “data245” data series has been identified, the proposed method is applied to “data300” data series. It requires the input data ( $v_{leader}(t)$ ,  $v_{follower}(t)$  and distance  $D(t)$ ) and exports the estimated  $v_{follower}(t + 0.5)$ . The RMSN values have been estimated per time instant  $t$  in order to compare predicted and observed speed values and estimate the performance of this methodological approach. The validation results are presented in Figs. 6, 7, 8 and 9.

In Fig. 6, the proposed method outperforms Gipps’ model and produces a more reliable speed prediction. The estimated RMSN for dataset “data300” is 0.19 using the Gipps’ model and 0.12 using the loess model. The flexible model outperforms the conventional model and produces a more reliable speed prediction.

In Figs. 7 and 8, an analysis of the results per vehicle type is attempted. Figure 7 shows the Empirical Cumulative Distribution Function (ECDF) of RMSN per vehicle type. The best performance of loess method is achieved for cars and light commercial vehicles, while higher RMSN are observed for other vehicle types, especially for trucks and auto-rickshaws. In Fig. 8 ECDF of RMSN are outlined per vehicle type of the leader when the follower is a car. Vehicles pairs car-car and motorcycle-car (leader-follower) have almost 80% of RMSN values lower than

**Table 1** Observations per vehicle type used for calibration of each data sample

a/a	Vehicle type	Sample 1	Sample 2	Sample 3
1	Motorcycle	2665	2701	2626
2	Car	1347	1292	1347
3	Bus	145	156	156
4	Truck	41	29	15
5	Light commercial vehicle	56	59	78
6	Auto-rickshaw	746	763	778



**Table 2** Initial values, bounds and optimization results for Gipps' parameter values

Parameters	Initial values	Constraints		Opt. values			Mean values
		Min	Max	Sample 1	Sample 2	Sample 3	
$\alpha$	0.8	0.8	-2.6	0.81	0.82	0.82	0.82
b	-3.2	-5.2	-1.6	-5.18	-5.17	-5.08	-5.14
V	14.4	10.4	29.6	10.45	10.44	10.44	10.44
s	5.9	5.6	7.5	5.62	5.60	5.60	5.61
b	-3.1	-4.5	-3.0	-3.01	-3.01	-3.00	-3.01
RMSN	-	-	-	0.21	0.22	0.21	-

0.1. The curve of vehicle pair truck–car corresponds to higher RMSN than the other vehicle pairs. It is evident that vehicle type plays a significant role in driving behavior.

Finally, in Fig. 9 observed speeds are plotted versus predicted speeds per vehicle type. Linearity is evident for all vehicle types.

**Identification of virtual lane changes**

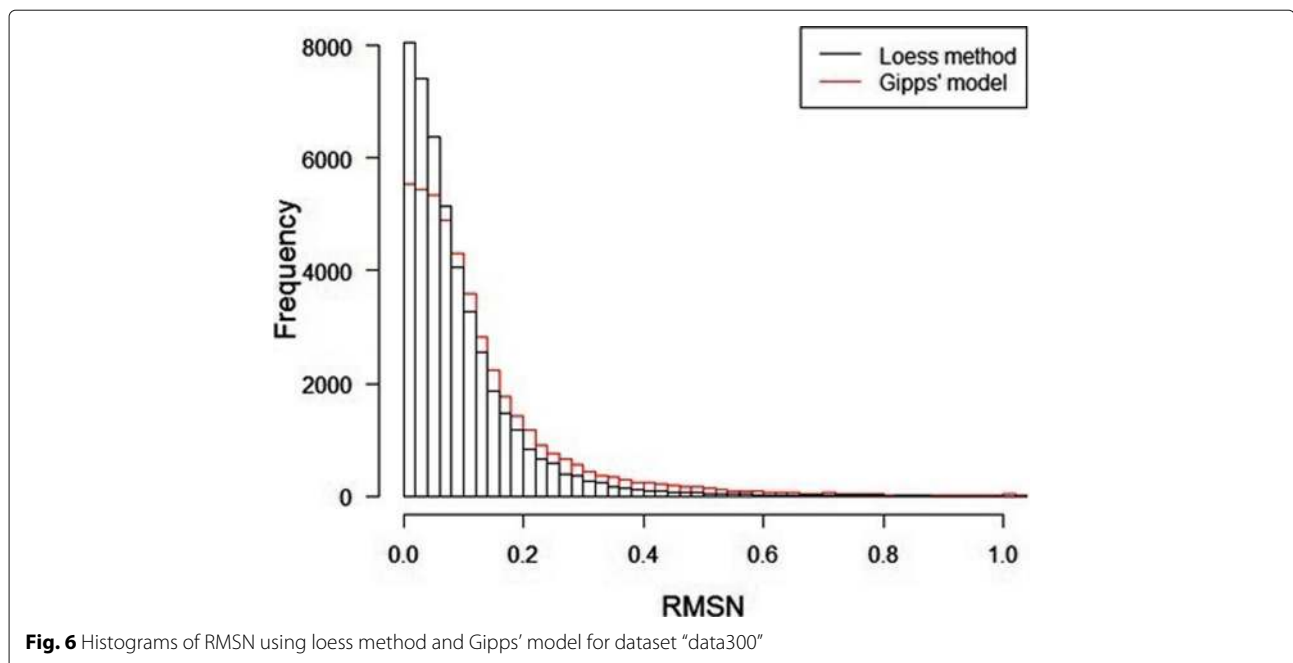
Models developed for lane-based traffic conditions may not be appropriate to simulate traffic situations in developing countries, where weak lane discipline is often observed. Traffic in the developing world is so heterogeneous that often lane-based models cannot be realistic. To overcome some of the associated limitations, in this research a methodology is proposed using temporary virtual lanes. An algorithm for the identification of significant lateral changes has been applied and the feasibility of the method has been explored.

**Breakpoints**

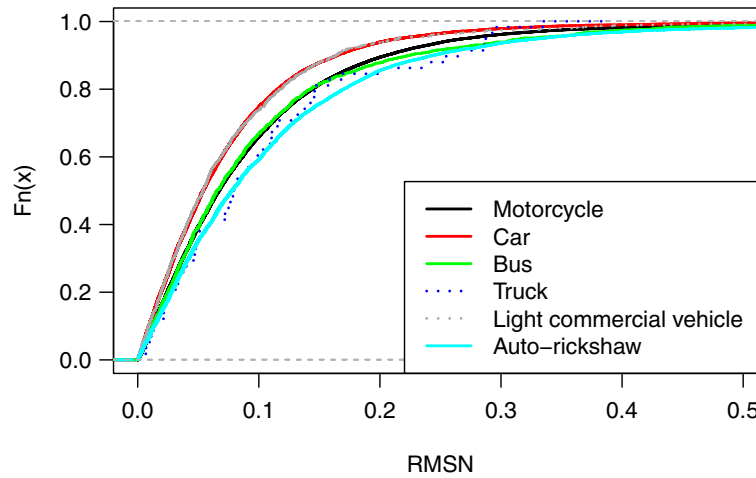
In order to identify structural changes in sequence of lateral positions, 'strucchange' package [42] was used in R statistical software [43]. This package is appropriate for testing, monitoring and dating structural changes in regression models. Breakpoints are marked in positions with significant lateral changes.

**Results**

The analysis is implemented using a few vehicles of the available datasets. It is mentioned that these vehicles are cars and their trajectories are extracted from dataset collected in the period 2:45–3:00 PM. In particular, the vehicle 109 is used for the first example. The optimal number of breakpoints is defined by the associated residual sum of squares (RSS) and Bayesian information criterion (BIC), as presented in Fig. 10a. Two breakpoints have been computed as the optimal breakpoints. To justify further the findings, F-statistics are estimated for the subject



**Fig. 6** Histograms of RMSN using loess method and Gipps' model for dataset "data300"



**Fig. 7** ECDF of RMSN per vehicle type for dataset “data300”

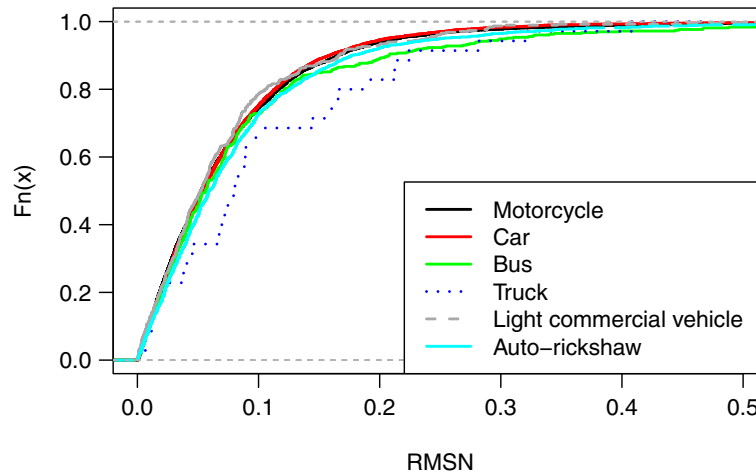
example and are plotted in Fig. 10b. The position of two breakpoints and the optimal segmentation of the data are indicated. It seems that identification of lane-changing manoeuvres is feasible. As it is observed in Fig. 10c, changes in lateral positions are small. This is attributed to the small vehicle size and small overlaps between the leader and the follower. Another example with vehicle 848 is also illustrated.

As far as vehicle 848 is concerned, breakpoints that are estimated by the algorithm are presented in Fig. 11. In Fig. 11c, the first breakpoint corresponds to a greater lateral movement than the second one.

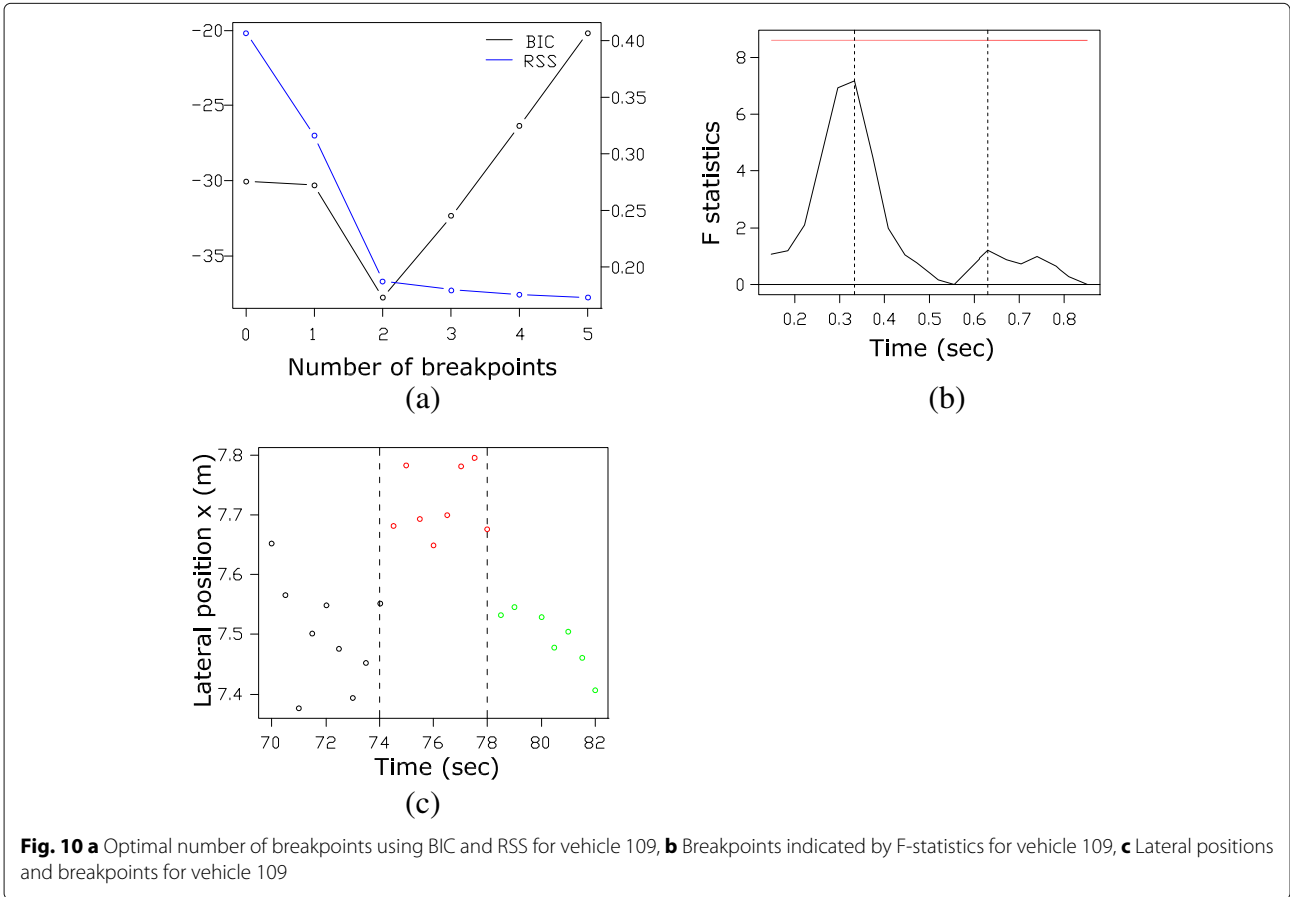
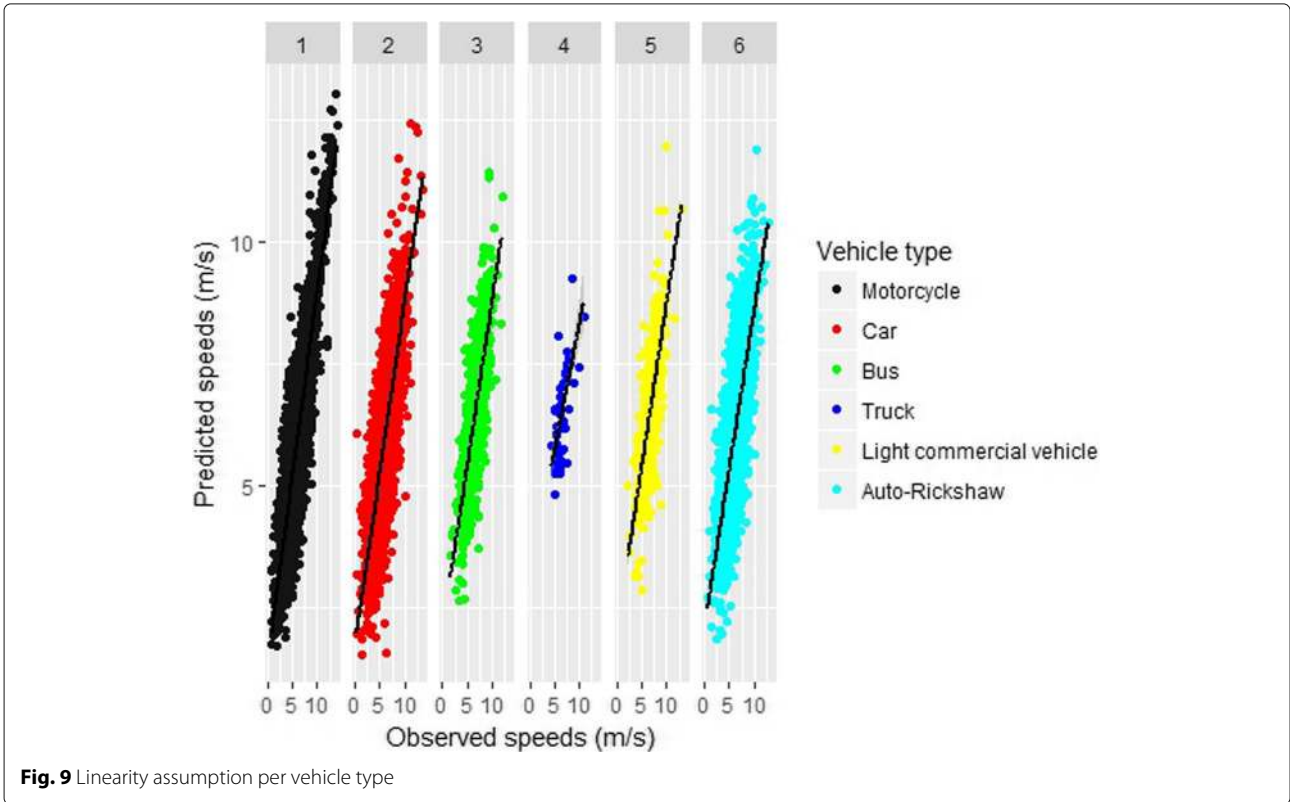
**Conclusions and future prospects**

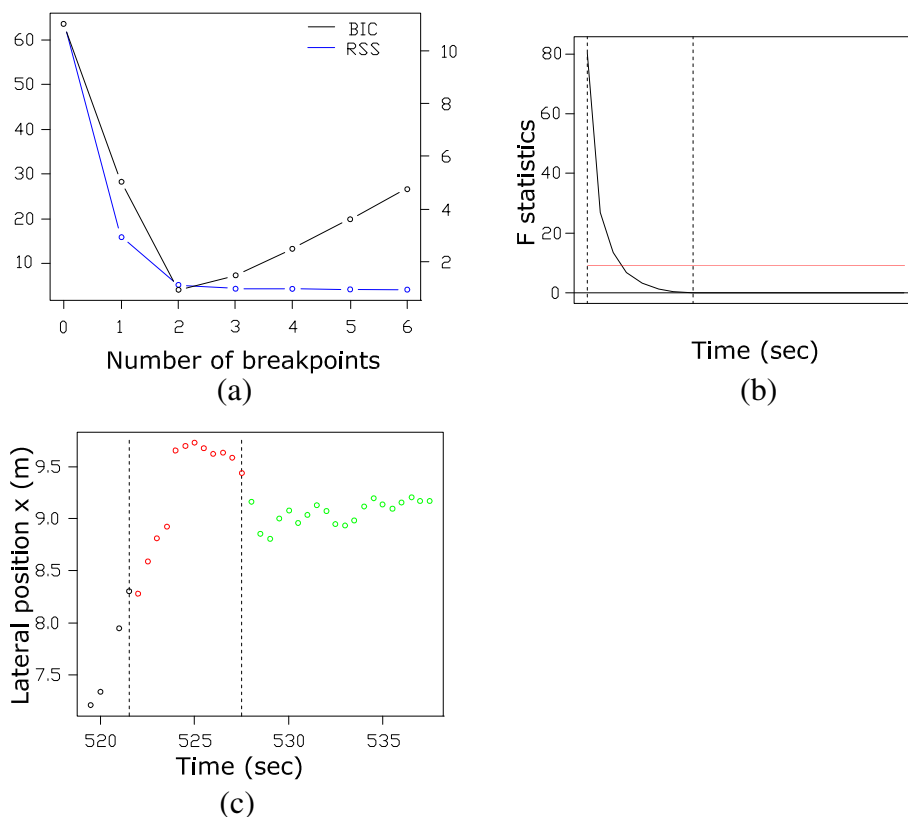
Models developed for lane-based traffic conditions may not be appropriate to simulate traffic situations in developing countries, where weak lane discipline is often

observed. Traffic in such conditions is so heterogeneous that often lane-based models cannot be realistic. To overcome these limitations, in this research a methodology is proposed based on data-driven models and using temporary virtual lanes. An algorithm for the identification of significant lateral changes has been applied and the feasibility of the method has been explored. In this research, the algorithm has identified all the breakpoints on the available data without constraints. However, the sensitivity of the algorithm could be further explored by setting a minimal segment size either given as fraction relative to the sample size or as an integer giving the minimal number of observations in each segment. The use of other algorithms, such as ‘segmented’ package [44] and ‘change-point’ [45], should be also checked for the same purpose. A method for estimation of virtual lane width has been also described.



**Fig. 8** ECDF of RMSN per vehicle type of the preceding vehicle when the follower is a car (dataset “data300”)





**Fig. 11** **a** Optimal number of breakpoints using BIC and RSS for vehicle 848, **b** Breakpoints indicated by F-statistics for vehicle 848, **c** Lateral positions and breakpoints for vehicle 848

Data driven approaches could be a promising tool for modeling mixed traffic. They lead to flexible car-following models and thus to more robust and reliable representation of driving behavior. This simple methodological approach outperforms the reference (Gipps') model for the available data. For the available data, speed prediction with RMSN 12% is achieved using loess method, while 19% using Gipps model. Data-driven estimation techniques are designed to address cases in which the traditional approaches do not perform well or cannot be effectively applied without including undue labor. Furthermore, the findings have interesting implications for the role of vehicle type. More specifically, vehicles pairs car-car and motorcycle-car (leader-follower) have almost 80% of RMSN values lower than 0.1, while the curve of vehicle pair truck-car corresponds to higher RMSN. Regarding the identification of lane-changing manoeuvres, breakpoints are marked in positions with significant lateral changes for few trajectories and seem to correspond to lane changing manoeuvres. However, further experimental analysis is required.

This research has highlighted the difficulties in modeling mixed traffic conditions and has explored the feasibility of data-driven models versus Gipps model in

this context. Different vehicle pairs resulted in different model efficiency, showing the need for vehicle-dependent models. Finally, this research contributed to the introduction of an alternative method for setting temporary virtual lanes under mixed traffic conditions. As the proposed methodology is data-driven, its transferability is feasible to any another section/ corridor or city in India and other developing countries. Furthermore, the proposed methodology is also useful for urban road networks without strict compliance to road traffic lanes, observed mainly in South European countries. More specifically, in Europe motorcycles and sometimes bicycles share the same road space with cars and tend to move through the lateral gaps. However, the appropriate input data should be used to fit the model for each case. It is suggested that training data come from similar network and traffic conditions with the explanatory data.

As future prospects, swarm-like models and crowd simulation models could also be considered for modeling mixed traffic and weak-lane discipline conditions. In addition, the proposed methodology allows incorporation of further variables moving towards an integrated solution for the simulation of mixed traffic. For instance, vehicle-dependent models need to be developed in case

of heterogeneous traffic, as the drivers of vehicles with unequal dimensions tend to have different driving behaviors; furthermore, different vehicle types are characterized by varying vehicle kinematics. Thus, it is foreseen that further exploration into this could open up opportunities to understand and simulate driving behavior in non-lane discipline conditions with heterogeneity of vehicle types.

#### Acknowledgments

The authors would like to thank Prof. Tomer Toledo from Technion - Israel Institute of Technology for making the data from India freely available.

#### Funding

There was no funding for this research. Not applicable.

#### Availability of data and materials

Data used in this research include weak lane-discipline trajectory data, which have been collected in India and are available at <http://toledo.net.technion.ac.il/downloads>. A detailed description of the data could be found in [30].

#### Authors' contributions

CA and VP developed the proposed methodology for modeling mixed traffic conditions. CA created the outline of the paper. VP validated the methodology using a case study and drafted the manuscript. All authors read and approved the final manuscript.

#### Authors' information

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### Author details

<sup>1</sup>School of Rural and Surveying Engineering, National Technical University of Athens, 9, Iroon Polytechniou St., Zografou Campus, 15780 Athens, Greece.

<sup>2</sup>Technical University of Munich, Arcisstrasse 21, 80333, Munich, Germany.

Received: 19 April 2018 Accepted: 4 December 2018

Published online: 29 December 2018

#### References

- Wong K, Lee T.-C., Chen Y.-Y. (2016) Traffic characteristics of mixed traffic flows in urban arterials. *Asian Transport Studies* 4(2):379–391
- Munigety C. R., Mathew T. V. (2016) Towards behavioral modeling of drivers in mixed traffic conditions. *Transportation in Developing Economies* 2(1):1–20
- Kanagaraj V., Asaithambi G., Kumar C. N., Srinivasan K. K., Sivanandan R. (2013) Evaluation of different vehicle following models under mixed traffic conditions. *Procedia-Social and Behavioral Sciences* 104:390–401
- Mathew T. V., Munigety C. R., Bajpai A. (2013) Strip-based approach for the simulation of mixed traffic conditions. *Journal of Computing in Civil Engineering* 29(5):04014069
- Asaithambi G., Kanagaraj V., Toledo T. (2016) Driving behaviors: Models and challenges for non-lane based mixed traffic. *Transportation in Developing Economies* 2(2):19
- Li Y., Zhang L., Peeta S., Pan H., Zheng T., Li Y., He X. (2015) Non-lane-discipline-based car-following model considering the effects of two-sided lateral gaps. *Nonlinear Dynamics* 80(1–2):227–238
- Parsuvanathan C. (2015) Proxy-lane algorithm for lane-based models to simulate mixed traffic flow conditions. *International Journal of Traffic and Transportation Engineering* 4(5):131–136
- Gundaliya P., Mathew T. V., Dhingra S. L. (2008) Heterogeneous traffic flow modelling for an arterial using grid based approach. *Journal of Advanced Transportation* 42(4):467–491
- Metkari M., Budhkar A., Maurya A. K. (2013) Development of simulation model for heterogeneous traffic with no lane discipline. *Procedia-Social and Behavioral Sciences* 104:360–369
- Choudhury C. F., Islam M. M. (2016) Modelling acceleration decisions in traffic streams with weak lane discipline: a latent leader approach. *Transportation research part C: emerging technologies* 67:214–226
- Maurya A. K. (2011) Comprehensive approach for modeling of traffic streams with no lane discipline. In: 2nd International Conference on Models and Technologies for Intelligent Transportation Systems
- Chunchu M., Kalaga R. R., Seethepalli N. V. S. K. (2010) Analysis of microscopic data under heterogeneous traffic conditions. *Transport* 25(3):262–268
- Lan L., Chang C. (2004) Motorcycle-following models of general motors (gm) and adaptive neuro-fuzzy inference system. *Transportation Planning Journal* 33(3):511–536
- Vasic J., Ruskin H. J. (2012) Cellular automata simulation of traffic including cars and bicycles. *Physica A: Statistical Mechanics and its Applications* 391(8):2720–2729
- Lee T.-C. (2007) An agent-based model to simulate motorcycle behaviour in mixed traffic flow. PhD thesis, Imperial College London (University of London)
- Lenorzer A., Casas J., Dinesh R., Zubair M., Sharma N., Dixit V., Torday A., Brackstone M. (2015) Modelling and simulation of mixed traffic. In: Australasian Transport Research Forum (ATRF), 37th, 2015, Sydney, New South Wales, Australia
- Liang X., Baohua M., Qi X. (2012) Psychological-physical force model for bicycle dynamics. *Journal of Transportation Systems Engineering and Information Technology* 12(2):91–97
- Chandra S. (2004) Capacity estimation procedure for two lane roads under mixed traffic conditions. *Journal of Indian Road Congress* 165:139–170
- Mehar A., Chandra S., Velmurugan S. (2014) Highway capacity through vissim calibrated for mixed traffic conditions. *KSCE journal of Civil Engineering* 18(2):639–645
- Papathanasopoulou V., Antoniou C. (2017) Flexible car-following models on mixed traffic trajectory data. In: Transportation Research Board 96th Annual Meeting
- Huval B., Wang T., Tandon S., Kiske J., Song W., Pazhayampallil J., Andriluka M., Cheng-Yue R., Mujica F., Coates A., Rajpurkar P., Migimatsu T., Y. Ng A. (2015) An empirical evaluation of deep learning on highway driving. arXiv preprint arXiv:1504.01716
- Chen X.-Y., Pao H.-K., Lee Y.-J. (2014) Efficient traffic speed forecasting based on massive heterogeneous historical data. In: Big Data (Big Data), 2014 IEEE International Conference On. IEEE, Washington, DC. pp 10–17
- Karlaftis M. G., Vlahogianni E. I. (2011) Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies* 19(3):387–399
- Mitchell T. M., et al. (1997) *Machine learning*. WCB. McGraw-Hill, Boston, MA
- Cleveland W. S. (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association* 74(368):829–836
- Cleveland W. S., Devlin S. J. (1988) Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association* 83(403):596–610
- Cohen R. A. (1999) An introduction to proc loess for local regression. In: Proceedings of the 24th SAS Users Group International Conference, Paper, vol. 273. Citeseer, North Carolina
- Cleveland W. S., Devlin S. J., Grosse E. (1988) Regression by local fitting: methods, properties, and computational algorithms. *Journal of econometrics* 37(1):87–114
- Antoniou C., Koutsopoulos H. N., Yannis G. (2013) Dynamic data-driven local traffic state estimation and prediction. *Transportation Research Part C: Emerging Technologies* 34:89–107
- Kanagaraj V., Asaithambi G., Toledo T., Lee T.-C. (2015) Trajectory data and flow characteristics of mixed traffic. *Transportation Research Record: Journal of the Transportation Research Board*:1–11
- Fritzsche H.-T. (1994) A model for traffic simulation. *Traffic Engineering+ Control* 35(5):317–21
- Siddharth S., Ramadurai G. (2013) Calibration of vissim for indian heterogeneous traffic conditions. *Procedia-Social and Behavioral Sciences* 104:380–389

33. Yulianto B. (2003) Application of fuzzy logic to traffic signal control under mixed traffic conditions. *Traffic Engineering and Control* 44(9):332–335
34. Van T. H., Schmoecker J.-D., Fujii S. (2009) Upgrading from motorbikes to cars: Simulation of current and future traffic conditions in ho chi minh city. In: *Proceedings of the Eastern Asia Society for Transportation Studies Vol. 7 (The 8th International Conference of Eastern Asia Society for Transportation Studies, 2009)*. Eastern Asia Society for Transportation Studies, Surabaya. pp 335–335
35. Olstam J. J., Tapani A. (2004) Comparison of car-following models. Technical report
36. Gipps P. G. (1981) A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological* 15(2):105–111
37. R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing. <https://www.R-project.org/>
38. Runarsson T. P., Yao X. (2005) Search biases in constrained evolutionary optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews* 35(2):233–243
39. Papathanasopoulou V., Antoniou C. (2015) Towards data-driven car-following models. *Transportation Research Part C: Emerging Technologies* 55:496–509
40. Sobol I. M. (2001) Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation* 55(1–3):271–280
41. Sfeir G., Antoniou C. (2017) Simulation-based evacuation planning using state-of-the-art sensitivity analysis techniques. Technical report
42. Zeileis A., Leisch F., Hornik K., Kleiber C. (2001) strucchange. an r package for testing for structural change in linear regression models
43. R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. R Foundation for Statistical Computing. <http://www.R-project.org/>
44. Muggeo V. M. (2008) Segmented: an r package to fit regression models with broken-line relationships. *R news* 8(1):20–25
45. Killick R., Eckley I. (2014) changepoint: An r package for changepoint analysis. *Journal of statistical software* 58(3):1–19

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)

---