

## Research Article

# A Support Vector Regression Approach for Investigating Multianticipative Driving Behavior

Bin Lu,<sup>1</sup> Shaoquan Ni,<sup>1</sup> and Scott S. Washburn<sup>2</sup>

<sup>1</sup>*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China*

<sup>2</sup>*Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL 32611, USA*

Correspondence should be addressed to Bin Lu; [binlu.transport@gmail.com](mailto:binlu.transport@gmail.com)

Received 22 January 2015; Revised 24 April 2015; Accepted 24 April 2015

Academic Editor: Valery Sbitnev

Copyright © 2015 Bin Lu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a Support Vector Regression (SVR) approach that can be applied to predict the multianticipative driving behavior using vehicle trajectory data. Building upon the SVR approach, a multianticipative car-following model is developed and enhanced in learning speed and prediction accuracy. The model training and validation are conducted by using the field trajectory data extracted from the Next Generation Simulation (NGSIM) project. During the model training and validation tests, the estimation results show that the SVR model performs as well as IDM model with respect to the model prediction accuracy. In addition, this paper performs a relative importance analysis to quantify the multianticipation in terms of the different stimuli to which drivers react in platoon car following. The analysis results confirm that drivers respond to the behavior of not only the immediate leading vehicle in front but also the second, third, and even fourth leading vehicles. Specifically, in congested traffic conditions, drivers are observed to be more sensitive to the relative speed than to the gap. These findings provide insight into multianticipative driving behavior and illustrate the necessity of taking into account multianticipative car-following model in microscopic traffic simulation.

## 1. Introduction

Microscopic traffic simulation tools are playing an increasingly essential role in Intelligent Transportation System (ITS). The problems in ITS application, such as adaptive congestion control, incident management, and traffic flow forecast, are usually difficult to solve due to the complex nature of the system dynamics in traffic flow. Microscopic traffic simulation tools permit the transportation engineers to study such problems under different scenarios without any disruption to real traffic. It is therefore of great importance to ensure that these simulation tools are able to replicate real traffic situations sufficiently well.

The accuracy of a simulation tool depends highly on its vehicle-movement models, especially the car-following model, which describes the driving behavior by which vehicles follow each other in the same lane. Much research has been done with regard to the car-following model since it was first proposed by Pipes [1]. The existing car-following models

can mainly be classified into six groups: (1) stimulus-response models (e.g., Gazis-Herman-Rothery (GHR) model [2]), (2) safe distance models (e.g., Gipps model [3]), (3) desired measures models (e.g., intelligent driver model (IDM) [4]), (4) cellular automata models (e.g., Nagel-Schreckenberg model [5]), (5) psychophysical models (e.g., Wiedemann model [6]), and (6) fuzzy logic-based models (e.g., Kikuchi and Chakroborty model [7]). More detailed reviews of these car-following models are available elsewhere [8–11].

Although numerous car-following models have been developed, inaccuracy and unreliability in many of these models' performance have been identified in earlier studies [12–17]. For instance, Brockfeld et al. [12] reported that minimum calibration errors of about 15% to 25% could probably not be suppressed, regardless of the model used in the experiments. They also inferred that the errors might result from a highly stochastic component in the driver's behavior. It turns out that better modeling approaches are needed to correctly explain the dynamic and stochastic driving behaviors.

One particular area that could lead to increased realism and accuracy in car-following models is that of multianticipation in follower's driving behavior. Researchers have indicated that driving behavior cannot be adequately described by only considering the immediate vehicle in front. Rather, drivers anticipate traffic conditions farther downstream by considering the second leading vehicle and even the third and fourth vehicles ahead [18–23]. Based on such an assumption, several well-known car-following models have been modified to incorporate multileader stimuli. More recently, Hoogendoorn et al. [24] proposed a generalized framework for multianticipative car-following model. Farhi et al. [25] and Zhang [26] developed multianticipative car-following models by using piecewise linear and multiple linear regression, respectively. These studies on multianticipative car-following models are of great importance due to the potential application in the context of microscopic traffic simulation, as most of the car-following models in commercial simulation programs consider only the first leader, whereas multiple leaders should be considered in the model to describe driving behavior more realistically.

Recently, with the advanced computing capability of processing high-resolution vehicle trajectory data, efforts have been made to improve the performance of car-following models by using Machine Learning (ML) methods (e.g., Artificial Neural Networks and Support Vector Regression (SVR)) [27–32]. The results of comparative experiments showed that ML-based models outperformed the traditional car-following models. In particular, Wei and Liu [32] reported that the complex asymmetric characteristic in car-following behavior could be well interpreted by using the SVR approach.

Motivated by the outstanding learning and generalization ability of the SVR approach, this study attempts to further explore the SVR-based car-following model, with a particular focus on the analysis of the multianticipation phenomenon in driving behavior, using high-resolution trajectory datasets (i.e., NGSIM [33]). The objective of this study is twofold: (1) to develop a multianticipative car-following model using the SVR approach and (2) to empirically investigate the number of vehicles and the type of stimuli to which drivers react.

## 2. Model Framework

In this section, we propose the multianticipative SVR-based car-following model after a brief discussion on the SVR approach. SVR was extended from Support Vector Machine to solve nonlinear regression problem [34]. It has been receiving increased attention because of its remarkable generalization performance, the absence of local minima, and the sparse representation of the solution. As SVR has the potential to model complex, real-world problems, it has been used in the study of several transportation topics [35–38].

The basic idea of SVR is to solve a nonlinear regression in a linear way by mapping nonlinear input data from original dimensional feature space to a higher dimensional feature space. Consider a set of training data as  $(x_1, y_1), \dots, (x_i, y_i), \dots, (x_N, y_N)$ , where  $x_i$  is the training sample,  $y_i$  is the corresponding target value, and  $N$  refers to

the number of the training data points. The SVR estimation function takes the following form:

$$f(x) = \omega^T \varphi(x) + b, \quad (1)$$

where  $\varphi(\cdot)$  denotes a nonlinear space transformation,  $\omega$  is the weight vector, and  $b$  is the bias. Our goal is to find a  $f(x)$  that has at most  $\varepsilon$  deviation from the target value  $y_i$ . One feasible way to achieve this goal is to solve the following convex optimization problem:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & \begin{cases} y_i - \omega^T \varphi(x_i) - b \leq \varepsilon + \xi_i \\ \omega^T \varphi(x_i) + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, \dots, N, \end{cases} \end{aligned} \quad (2)$$

where  $C$  is a predefined penalty,  $\varepsilon$  is the tolerance, and  $\xi_i, \xi_i^*$  are the positive and negative slack variables, respectively. Applying Lagrange multipliers and Karush-Kuhn-Tucker (KKT) condition to (2) yields the following dual Lagrange dual function:

$$\begin{aligned} \max \quad & -\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \varphi(x_i)^T \varphi(x_j) \\ & - \varepsilon \sum_{i=1}^N (\alpha_i + \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C], \end{cases} \end{aligned} \quad (3)$$

where  $\alpha_i$  and  $\alpha_i^*$  are Lagrange multipliers. In deriving (3), the partial derivative of Lagrange function with respect to the weight vector  $\omega$  has to be eliminated for optimality:

$$\omega - \sum_{i=1}^N (\alpha_i - \alpha_i^*) x_i = 0. \quad (4)$$

By further introducing a kernel function,  $\omega$  is no longer needed to be given explicitly in (1). The SVR estimation function becomes

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) k(x, x_i) + b, \quad (5)$$

where  $k(x, x_i) = \varphi(x_i)^T \varphi(x)$  is the kernel function. The kernel function provides a linear classifier to solve nonlinear problem in a computationally efficient way. The most commonly used kernel functions are Linear, Polynomial, Sigmoid, and Radial Basis Function (RBF). Though no general guideline is available to choose a kernel function, the RBF kernel is generally preferred because of fewer parameters included and less numerical difficulties [39]; thus, it is employed in

this study for experimentation. The RBF kernel is defined as follows:

$$k(x, x_i) = \exp(-\gamma \|x - x_i\|^2), \quad (6)$$

where  $\gamma$  is the kernel width tuning the prediction accuracy. Since (2) is a convex optimization problem, there always exists a unique solution for  $\omega$  and  $b$  once the kernel function along with the parameters  $C$ ,  $\epsilon$ , and  $\gamma$  is determined. The reader is referred to [40] for further explanation.

To improve learning speed and prediction quality, it is also necessary to perform feature selection to discard redundant or irrelevant variables while only extracting a minimum set of indispensable variables from the input data [41]. It is known that vehicle speed, relative speed, and gap are widely considered the most decisive stimuli in the car-following models (e.g., GHR and IDM) and most importantly can be directly observed from vehicle trajectory data. Therefore, we can define the SVR-based car-following model as the following acceleration function:

$$\begin{aligned} a(t) = f(v(t), \Delta v(t - \tau), \Delta s(t - \tau)) &= \sum_{i=1}^N (\alpha_i - \alpha_i^*) \\ &\cdot k\{[v_i(t), \Delta v_i(t - \tau), \Delta s_i(t - \tau)], \\ &[v(t), \Delta v(t - \tau), \Delta s(t - \tau)]\} + b, \end{aligned} \quad (7)$$

where  $a(t)$  is the dependent variable representing the acceleration of following vehicle at time  $t$ ,  $v(t)$  is the vehicle speed,  $\Delta v(t - \tau)$  denotes the relative speed, and  $\Delta s(t - \tau)$  denotes the gap (i.e., the distance between the back bumper of the leader and the front bumper of the follower) at a previous time step  $t - \tau$ ;  $\tau$  is the driver's reaction time. Initially, we build the SVR-based car-following model to consider only one leading vehicle. It is found in the literature that multianticipative car-following models are often extensions of the single-leader model [24]. Thus, (7) is further modified to account for multianticipation by incorporating leader-specific parameters:

$$\begin{aligned} a(t) = f(v(t), \Delta v^{(1)}(t - \tau), \Delta v^{(2)}(t - \tau), \dots, \Delta v^{(p)}(t - \tau), \Delta s^{(1)}(t - \tau), \Delta s^{(2)}(t - \tau), \dots, \Delta s^{(p)}(t - \tau)) \\ = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \\ \cdot k \left\{ \begin{bmatrix} v_i(t), \\ \Delta v_i^{(1)}(t - \tau), \Delta v_i^{(2)}(t - \tau), \dots, \Delta v_i^{(p)}(t - \tau), \\ \Delta s_i^{(1)}(t - \tau), \Delta s_i^{(2)}(t - \tau), \dots, \Delta s_i^{(p)}(t - \tau) \end{bmatrix}, \right. \\ \left. \begin{bmatrix} v(t), \\ \Delta v^{(1)}(t - \tau), \Delta v^{(2)}(t - \tau), \dots, \Delta v^{(p)}(t - \tau), \\ \Delta s^{(1)}(t - \tau), \Delta s^{(2)}(t - \tau), \dots, \Delta s^{(p)}(t - \tau) \end{bmatrix} \right\} + b, \end{aligned} \quad (8)$$

where  $p \geq 1$  represents the number of leaders to whose relative speed and gap a driver reacts. This model is referred

to as SVR- $p$  in the rest of the paper. For example, SVR-3 model assumes a driver responding to the first, second, and third leaders. So far, the SVR-based car-following model is established and expected to deal with the multianticipative driving behavior.

### 3. Data Preparation

This study uses the Next Generation Simulation (NGSIM) trajectory dataset, which was collected on a 500-m segment of eastbound I-80 in Emeryville, California, from 5:00 to 5:15 p.m. on April 13, 2005, with 0.1-s resolution. It should be mentioned that the trajectory datasets were collected during peak hours that showed congestion. A more detailed introduction to this trajectory dataset is presented in NGSIM Interstate 80 Freeway Dataset Report [33]. To investigate multianticipative driving behavior, the platoons with a minimum size of five vehicles are considered in this study. Moreover, this study considers the following criteria to choose the appropriate car-following platoons.

- (i) The car-following platoons should be observed consecutively for at least a period of 60 s (i.e., 600 observations) to ensure sufficient observations for SVR learning.
- (ii) The car-following platoons present on Lane 1 (high occupancy lane) and Lane 7 (on-ramp) should be excluded.
- (iii) Only the car-following platoons without lane changing for any of its member during the observation period are accepted.
- (iv) The vehicle type of both followers and leaders in the platoon should only be passenger car. Considering the potentially different driving behavior from other vehicle types in the analysis is beyond the scope of this study.

VTAPE [42], a vehicle trajectory analysis tool developed by the authors, is used throughout this study to process NGSIM dataset. By applying the criteria proposed above, a total of 229 car-following platoons consisting of 202,598 trajectory points are extracted from the NGSIM dataset.

Note that the raw trajectory dataset, particularly the vehicle speed, has been denoised by using a symmetric exponential moving average filter, since studies have reported issues in the quality of the NGSIM dataset [43–45]. For a visual inspection, a sample of trajectories, speed and gap profiles from a car-following platoon are presented in Figure 1. The trajectory data of this platoon will also be used as the example in the next section for model performance comparison.

Before model training and validation, we also normalized the trajectory data to improve the SVR training convergence rate. The data normalization used in this study is to map the raw trajectory data into range  $[0, 1]$ :

$$x_{i,\text{norm}} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (9)$$

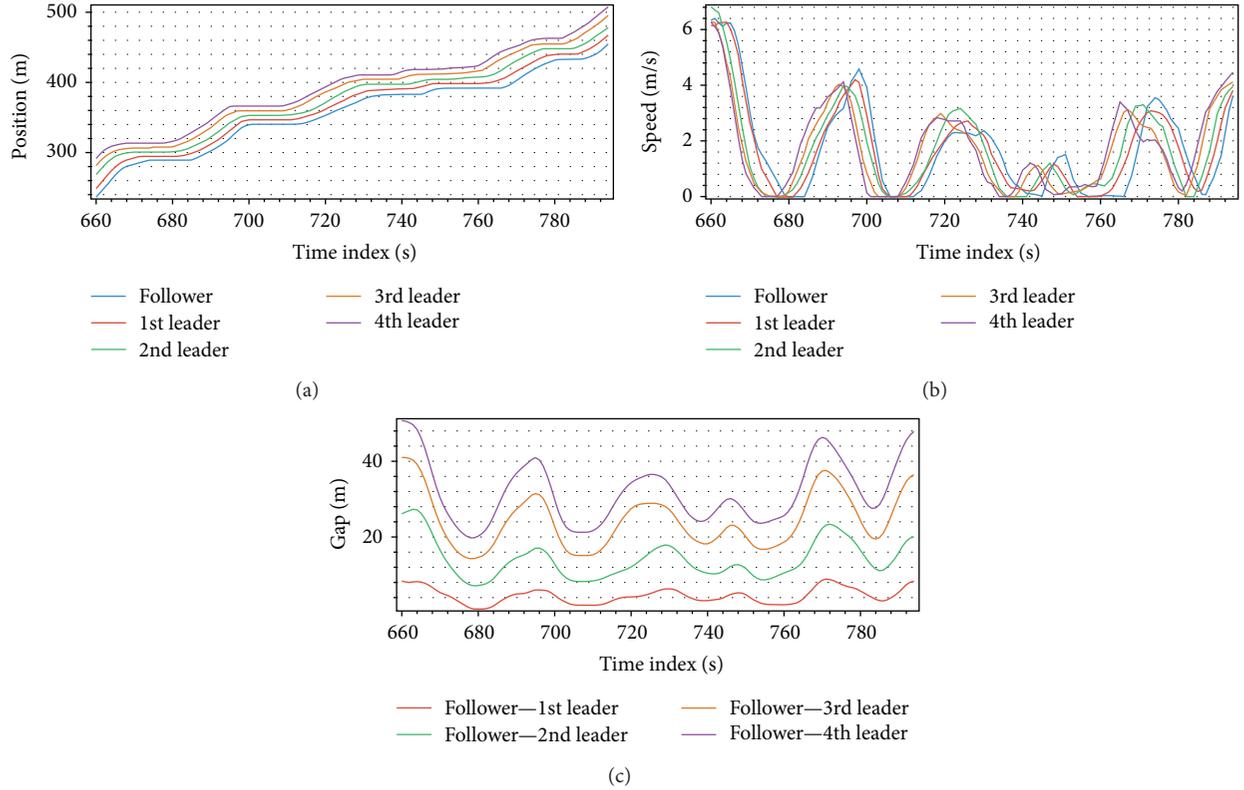


FIGURE 1: (a) Platoon trajectories; (b) platoon speed profiles; (c) platoon gap profiles.

where  $x_{i,\text{norm}}$  denotes the normalized value of  $x_i$ ,  $x_{\text{max}}$  is the maximum value of the training data, and  $x_{\text{min}}$  is the minimum value of the training data. The output value  $x_{i,\text{norm}}$  will be denormalized before comparing with its target value.

#### 4. Model Training and Validation

The objective of model training is to determine the set of SVR parameters that best fit the given data samples. In this paper, recall that for (8) we need to calibrate penalty  $C$ , tolerance  $\epsilon$ , and kernel width  $\gamma$  by vehicle speed  $v$ , relative speed  $\Delta v$ , gap  $\Delta s$ , and driver's reaction time  $\tau$ . However, the driver's reaction time is not directly observable from trajectory data. Efforts to estimate the reaction time have been made for a single-leader car-following model [29, 31, 46], while a systematic approach to estimate reaction time for multianticipation is still absent. In this study, we treat the reaction time as one of the model parameters to be calibrated. The feasible range for the reaction time is set from 0.5 to 2 s and discretized by the step of 0.1 s. Accordingly, the reaction time and other parameters that yield the best estimation result are considered the optimal parameters. The initial selection of remaining parameters is set as  $C = 2^{-5}, 2^{-3}, 2^{-1}, \dots, 2^{15}$ ,  $\epsilon = 0.001, 0.01, 0.1, 0.2, 0.5$ ,  $\gamma = 2^{-5}, 2^{-4}, 2^{-3}, \dots, 2^1$  and then optimized based on the cross-validation, which is a widely used trial-and-error testing for assessing the model

generalization ability. The idea of the cross-validation is to train on one subset of the samples and then to test the prediction accuracy on the other separate subsets. Although the cross-validation leads to high prediction accuracy, the implementation of the cross-validation could be very time-consuming on the large-scale data, since a full retraining of data samples should be done at each iteration. To reduce computational burden, the Accurate Online SVR (AOSVR) algorithm proposed by Ma et al. [47] is used in this study. AOSVR is able to update the trained SVR function whenever a sample is added or removed from the original training set, without the need to retrain from scratch. In this manner, during the cross-validation procedure, we first construct the SVR function using the entire trajectory dataset, followed by removing each subset (in our case, the trajectory data of each platoon) from the SVR function using the AOSVR algorithm, and then calculate the prediction error for each subset. Ultimately, we choose the parameters associated with the minimum error presented by an error indicator. The root mean square error (RMSE) is one of the basic error indicators, which measures the deviation of the predicted value from the field data:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}, \quad (10)$$

where  $\hat{y}_i$  presents the estimation of field data  $y_i$ . Another error indicator is Theil's inequality coefficient [48], widely used in econometrics, which is defined as

$$U = \frac{\sqrt{(1/N) \sum_{i=1}^N (\hat{y}_i - y_i)^2}}{\sqrt{(1/N) \sum_{i=1}^N \hat{y}_i^2} + \sqrt{(1/N) \sum_{i=1}^N y_i^2}}. \quad (11)$$

Theil's inequality coefficient provides a measurement of RMSE in relative terms, which falls between 0 and 1. The closer the value of  $U$  to zero, the better the prediction result.

Basically, either speed or gap can be used as the error term. Ossen [23] reported that the multicriterion objective function containing both speed and gap produced better results than either single-objective function. Thus, the following error indicator with multicriterion objective is applied in this study:

$$U^* = \frac{1}{2} \frac{\sqrt{(1/N) \sum_{i=1}^N (\hat{v}_i - v_i)^2}}{\sqrt{(1/N) \sum_{i=1}^N \hat{v}_i^2} + \sqrt{(1/N) \sum_{i=1}^N v_i^2}} + \frac{1}{2} \frac{\sqrt{(1/N) \sum_{i=1}^N (\Delta \hat{s}_i - \Delta s_i)^2}}{\sqrt{(1/N) \sum_{i=1}^N \Delta \hat{s}_i^2} + \sqrt{(1/N) \sum_{i=1}^N \Delta s_i^2}}, \quad (12)$$

where  $\hat{v}_i$  and  $\Delta \hat{s}_i$  are the estimation of observed speed and observed gap, respectively.

Note that the car-following behavior relies heavily on its preceding state, which means that the trajectory data cannot be regarded as independent samples. Thus, a car-following simulation is carried out in this study for model performance evaluation. During the simulation, the following vehicle starts with the same speed and position as its first observation, while the movement of leaders will be simulated exactly the same as in the observations. At each time step, the considered car-following model is applied to update the followers' trajectories according to (13). The error indicator (12) can be calculated at the end of the simulation:

$$\begin{aligned} v(t + \Delta t) &= v(t) + a(t) \Delta t \\ s(t + \Delta t) &= x(t) + v(t) \Delta t + \frac{1}{2} a(t) (\Delta t)^2, \end{aligned} \quad (13)$$

where  $s(t)$  is the position of the subject vehicle at time  $t$  and  $\Delta t$  is the time step (0.1 s).

In addition, to assess the performance of SVR model to conventional physical models, IDM model, which is considered one of the most robust car-following models in the literature, is applied to the trajectory data based on the same evaluation and simulation procedure. The IDM is defined as follows:

$$\begin{aligned} a(t) &= f(v(t), \Delta v(t), \Delta s(t)) \\ &= a^* \left[ 1 - \left( \frac{v(t)}{v^*} \right)^4 - \left( \frac{\Delta s^*(v(t), \Delta v(t))}{\Delta s(t)} \right)^2 \right] \end{aligned} \quad (14)$$

with the minimum desired gap  $\Delta s^*$  given by

$$\Delta s^*(v(t), \Delta v(t)) = \Delta s_0 + T_0 v(t) + \frac{v(t) \Delta v(t)}{2\sqrt{a^* b^*}}, \quad (15)$$

TABLE 1: Optimal parameters and training results for SVR models.

Parameter	SVR-1	SVR-2	SVR-3	SVR-4
$C$	4	4	2	2
$\epsilon$	0.1	0.1	0.1	0.1
$\gamma$	0.5	0.5	0.25	0.25
$\tau$ [s]	1.0	1.1	1.1	1.1
$U^*$	0.1430	0.1384	0.1362	0.1360

where  $v^*$ ,  $a^*$ , and  $b^*$  are the desired speed, maximum desired acceleration, and absolute maximum desired deceleration of the following vehicle, respectively;  $\Delta s_0$  is the jam distance;  $T_0$  denotes the desired time gap. By introducing a set of weighting coefficients, IDM can be extended to a linear combination form that accounts for the multianticipation driving behavior [49]:

$$\begin{aligned} a(t) &= \sum_{k=1}^p \lambda_k f(v(t), \Delta v(t), \Delta s(t)) \\ &= \sum_{k=1}^p \lambda_k a^* \left[ 1 - \left( \frac{v(t)}{v^*} \right)^4 - \left( \frac{\Delta s^*(v(t), \Delta v(t))}{\Delta s(t)} \right)^2 \right] \end{aligned} \quad (16)$$

$$\text{s.t. } 0 \leq \lambda_k \leq 1, \quad \sum_{k=1}^p \lambda_k = 1,$$

where weighting coefficients  $\lambda_k$  ( $k = 1, 2, \dots, p$ ) are assumed to decrease monotonically as  $k$  increases, which indicates the influence order of the leading vehicles and is in agreement with daily driving experience. This model is referred to as IDM- $p$  in the rest of the paper. To limit the parameter space for the optimization, the following constraints are applied to the minimum and maximum values of model parameters by referring to the study of Kesting and Treiber [50]. The desired speed  $v^*$  is restricted to [1, 70] m/s, the desired time gap  $T_0$  to [0.1, 5] s, and the jam distance  $\Delta s_0$  to [0.1, 8] m and both the maximum acceleration  $a^*$  and deceleration  $b^*$  are restricted to [0.1, 6] m/s<sup>2</sup>.

This study applies the first 12-min trajectory data, which covers 163 car-following platoons, to train the total of eight car-following models considered (i.e., SVR-1, SVR-2, SVR-3, and SVR-4; IDM-1, IDM-2, IDM-3, and IDM-4). By applying the described calibration method, we found the best fit of the car-following models to the training trajectory dataset. Tables 1 and 2 list the optimal parameters and training results for the SVR and IDM models, respectively.

Theil's inequality coefficients presented in Tables 1 and 2 are very close to zero, in the range of 0.1360 to 0.1430, which indicate that both the SVR and IDM models in the simulation tests perform well and are able to produce trajectories that closely match the field observations. It is also obvious that the models including multiple leading vehicles produce higher prediction accuracy, as corresponding Theil's inequality coefficients are smaller compared with the single-leader model. These training results imply that drivers may react not only to the first leader but also to the second, third, and even the fourth leaders. In addition, by comparing Theil's inequality coefficients of SVR and IDM models, we can notice

TABLE 2: Optimal parameters and training results for IDM models.

Parameter	IDM-1	IDM-2	IDM-3	IDM-4
$v^*$ [m/s]	24.00	24.00	24.00	24.00
$a^*$ [m/s <sup>2</sup> ]	1.02	1.00	1.00	0.96
$b^*$ [m/s <sup>2</sup> ]	3.13	2.95	2.96	2.92
$\Delta s_0$ [m]	2.73	2.61	2.42	2.33
$T_0$ [s]	1.38	1.41	1.33	1.36
$\lambda_1$	1.00	0.72	0.68	0.65
$\lambda_2$	—	0.28	0.20	0.18
$\lambda_3$	—	—	0.12	0.10
$\lambda_4$	—	—	—	0.07
$U^*$	0.1425	0.1399	0.1388	0.1371

TABLE 3: Validation results for SVR and IDM models.

Model	Number of best models	% best models	$U^*$
SVR-1	5	7.6	0.1514
SVR-2	8	12.1	0.1477
SVR-3	10	15.2	0.1458
SVR-4	11	16.7	0.1452
IDM-1	6	9.1	0.1511
IDM-2	7	10.6	0.1483
IDM-3	9	13.6	0.1472
IDM-4	10	15.2	0.1469

that the SVR models perform as well as the IDM models. But the consistency of the model performance needs to be determined by additional validation test.

Therefore, to further verify the models' performance in terms of predicting multianticipation, this study uses the remaining 3-min trajectory data, including 66 car-following platoons, to perform the validation test. Similar to the model training, the SVR and IDM models are evaluated in a microscopic simulation while applying the same optimal parameters derived from the model training results. Additionally, the number of times a specific model outperforms the other models in terms of prediction accuracy is counted during the simulation. Table 3 presents the overview of the validation results.

As might have been expected, the prediction errors shown in Table 3 tend to be slightly higher than in the model training with additional 6% error rate. But the range of Theil's inequality coefficients still falls in 0.1452 to 0.1514, indicating overall acceptable validation results, which is consistent with the typical error ranges reported in previous studies [12]. Recall that these eight car-following models are assumed to represent different driving styles according to the different number of leaders considered. Thus, it can be inferred that the model outperforming all the other models is most likely the driving style applied by the driver. From Table 3, we can see that 83.3% of the best models consider two or more leaders; even more than 60.6% are the three and four leader models. This result empirically indicates that drivers react to the behavior of not only the immediate leader in front but also the second and even the third and fourth leaders ahead. In general, the models including more leaders yield better

performance. However, as far as the model prediction accuracy is concerned, the result shows that there is still no single model that consistently outperforms all other models in all conditions. It turns out that different car-following models are still needed to accurately predict different multianticipative driving behaviors. Figure 2 compares the deviations of the speed and gap profiles generated by SVR and IDM models with a different number of leading vehicles considered (the original empirical trajectories were shown in Figure 1).

In order to quantitatively determine to what extent drivers react to the stimuli during multianticipative car following, a relative importance analysis is performed in the next section.

## 5. Relative Importance Analysis

Compared with the conventional statistical models (e.g., linear regression) in which the relative importance of each independent variable can be clearly explained by its coefficients, variable relative importance for SVR is difficult to determine due to the inherent limited input-output model framework. This study applies a derivative-based sensitivity analysis to measure the relative importance of features (independent variables) for SVR. For further explanation, we refer to the original work of Cao et al. [51].

Recall that only the data points with a nonzero coefficient, which are the so-called support vectors, are presented in (5). Hence, (5) can be rewritten as

$$f(x_i) = \sum_{j=1}^{N_S} (\alpha_j - \alpha_j^*) k(x_i, x_j) + b, \quad (17)$$

where  $N_S$  is the number of support vectors. The importance of support vector associated with the  $m$ th feature can be derived by calculating the partial derivative of (17) as follows:

$$\begin{aligned} \frac{\partial f(x_i)}{\partial x_{i,m}} &= \frac{\partial \sum_{j=1}^{N_S} (\alpha_j - \alpha_j^*) k(x_i, x_j)}{\partial x_{i,m}} + \frac{\partial b}{\partial x_{i,m}} \\ &\doteq \sum_{j=1}^{N_S} (\alpha_j - \alpha_j^*) \frac{\partial k(x_i, x_j)}{\partial x_{i,m}}. \end{aligned} \quad (18)$$

Substituting the RBF kernel function (6) into (18) yields

$$\begin{aligned} \frac{\partial f(x_i)}{\partial x_{i,m}} &= -2\gamma \sum_{j=1}^{N_S} (\alpha_j - \alpha_j^*) (x_i - x_j) \exp(-\gamma \|x_i - x_j\|^2). \end{aligned} \quad (19)$$

The average absolute deviation is one of the applicable measurements to evaluate the feature importance, which takes the following form:

$$S_m = \frac{\sum_{i=1}^N |\partial f(x_i) / \partial x_{i,m}|}{N}. \quad (20)$$

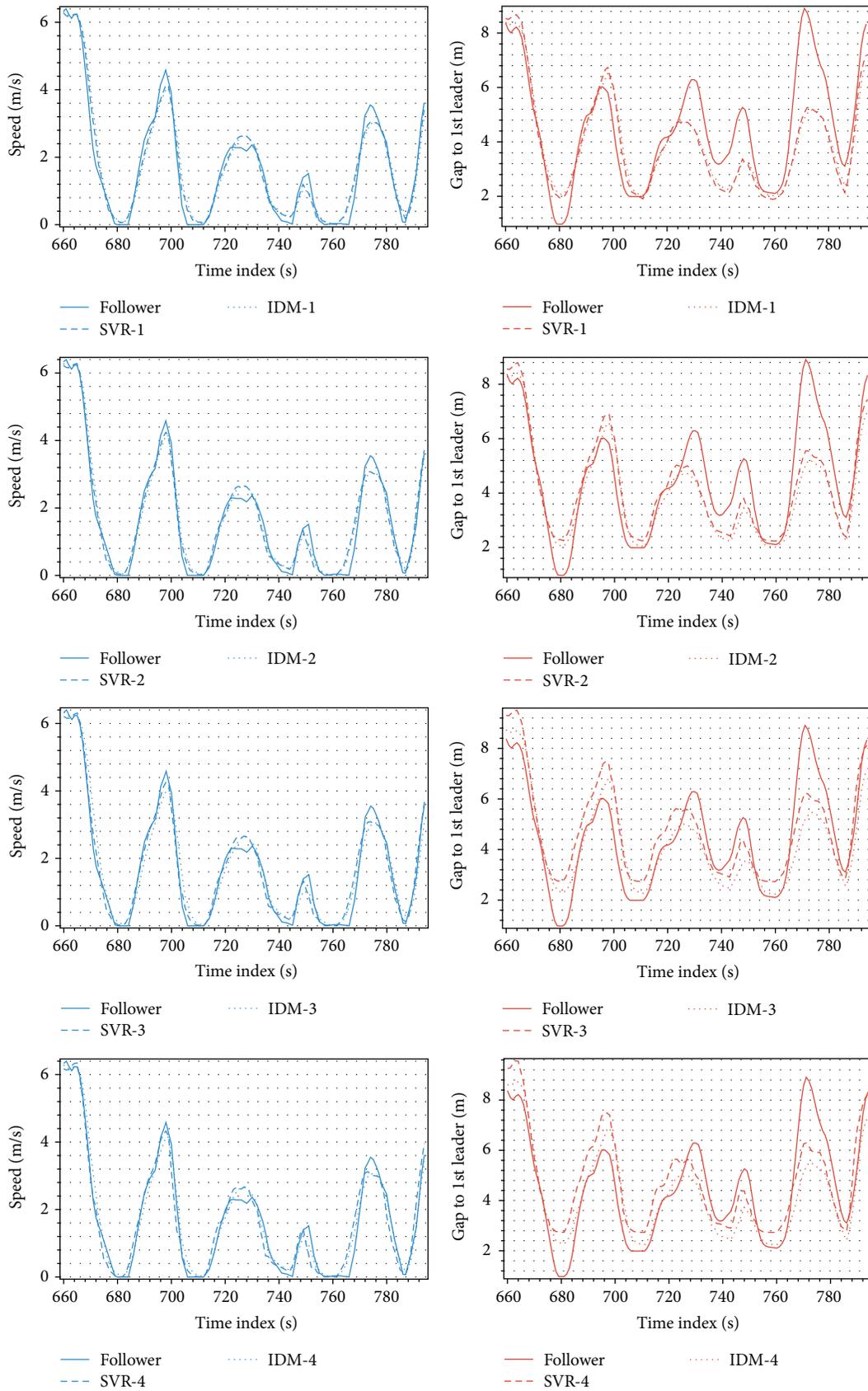


FIGURE 2: Comparison of speed and gap profiles generated by SVR and IDM models.

TABLE 4: Results of relative importance analysis.

Model	$\nu$	$\Delta\nu^{(1)}$	$\Delta s^{(1)}$	$\Delta\nu^{(2)}$	$\Delta s^{(2)}$	$\Delta\nu^{(3)}$	$\Delta s^{(3)}$	$\Delta\nu^{(4)}$	$\Delta s^{(4)}$
SVR-1	0.2105	0.5018	0.2877	—	—	—	—	—	—
SVR-2	0.1613	0.3844	0.2204	0.1492	0.0847	—	—	—	—
SVR-3	0.1384	0.3299	0.1891	0.1280	0.0728	0.0830	0.0588	—	—
SVR-4	0.1245	0.2967	0.1701	0.1151	0.0653	0.0747	0.0531	0.0684	0.0321

Based on (20), the relative importance for each feature can then be determined:

$$\begin{aligned}
 S_m^* &= \frac{S_m}{\sum_{m=1}^{N_F} S_m} \\
 &= \frac{\sum_{i=1}^N \left| -2\gamma \sum_{j=1}^{N_s} (\alpha_j - \alpha_j^*) (x_i - x_j) \exp(-\gamma \|x_i - x_j\|^2) \right|}{\sum_{m=1}^{N_F} \sum_{i=1}^N \left| -2\gamma \sum_{j=1}^{N_s} (\alpha_j - \alpha_j^*) (x_i - x_j) \exp(-\gamma \|x_i - x_j\|^2) \right|}
 \end{aligned} \quad (21)$$

where  $N_F$  is the number of features. To identify to what extent the different stimuli contribute in motivating the multianticipative driving behavior, we perform the proposed relative importance analysis to evaluate the nine multianticipative car-following models. Table 4 shows the importance scores for each model feature.

The importance scores presented in Table 4 are assumed to represent the sensitivity at which drivers react to the stimuli. It is apparent from Table 4 that drivers are overall most sensitive to the first leader. Not much difference is found between the sensitivity to the third leader and the sensitivity to the fourth leader. This may explain that the models considering four leaders do not yield much better prediction accuracy than the models considering three leaders. Such result is consistent with our daily driving experiences, as drivers may not be able to pay attention to four or more leading vehicles ahead. Additionally, combined with the finding in the model training and validation, it can be concluded that considering second, third, and even fourth leaders will further improve the model performance over a single-leader model. It is also evident that drivers appear to be more responsive to the relative speed than to the gap in this analysis. Since the trajectory data used for model training and validation was collected in an afternoon peak hour, the multianticipative driving behavior can in part be explained by the traffic congestion, as intuition suggests that drivers in congestion may react to the stimuli of multiple leaders for the sake of safety.

## 6. Summary and Future Research

The motivation for this study was to investigate multianticipative car-following behavior and provide empirical evidence for the interpretation. Efforts were made to build the SVR-based car-following model that considers multiple leaders. Due to its outstanding learning ability, the SVR-based car-following model is able to track the multianticipative driving behavior from the NGSIM trajectory dataset and performs as well as the IDM model. Since the IDM model has difficulty in

determining the model parameters (e.g., desired speed and desired gap) that are unobservable in nature, SVR model has the advantage that it provides a data-driven way that allows us to empirically study driving behavior directly from field trajectory data. The estimation results from training and validation tests using SVR and IDM models show that incorporating multianticipative behavior considerably improves the model prediction accuracy. Although both SVR and IDM models perform well in the validation test, this study shows that there is still no single model that consistently outperforms all the other models in all conditions. Different car-following models are required to accurately predict driving behaviors for different drivers. The validation results confirm that drivers react to the behavior of not only the immediate leader in front but also the second, third, and even more leaders. Additionally, this study performs a relative importance analysis to quantify the different stimuli to which drivers react in multileader car following. It is evident that drivers appear to be more sensitive to the relative speed than to the gap in congested traffic conditions.

These findings are of great importance for the current microscopic simulation practice. Most of the commercial simulation models include only the first leader; however, multiple leaders need to be considered to describe driving behavior more realistically. Future research will be aimed at extending the current empirical analysis to different levels of traffic flow and examining how the current results might change. Furthermore, applying various Machine Learning techniques to interpret complex traffic phenomenon and its implementation in the traffic simulation deserves more attention.

## Conflict of Interests

The authors declare that there is no conflict of commercial or associative interests regarding the publication of this work.

## Acknowledgment

The authors are grateful to the USA Federal Highway Administration for providing the NGSIM trajectory dataset.

## References

- [1] L. A. Pipes, "An operational analysis of traffic dynamics," *Journal of Applied Physics*, vol. 24, no. 3, pp. 274–281, 1953.
- [2] D. C. Gazis, R. Herman, and R. W. Rothery, "Nonlinear follow-the-leader models of traffic flow," *Operations Research*, vol. 9, no. 4, pp. 545–567, 1961.

- [3] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation Research Part B: Methodological*, vol. 15, no. 2, pp. 105–111, 1981.
- [4] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical Review E*, vol. 62, no. 2B, pp. 1805–1824, 2000.
- [5] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," *Journal de Physique I*, vol. 2, no. 12, pp. 2221–2229, 1992.
- [6] W. Leutzbach and R. Wiedemann, "Development and applications of traffic simulation models at the karlsruhe institut fur verkehrswesen," *Traffic Engineering & Control*, vol. 27, no. 5, pp. 270–275, 1986.
- [7] S. Kikuchi and P. Chakroborty, "Car-following model based on fuzzy inference system," *Transportation Research Record* 1365, Transportation Research Board, 1992.
- [8] M. Brackstone and M. McDonald, "Car-following: a historical review," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 2, no. 4, pp. 181–196, 1999.
- [9] T. Toledo, "Driving behaviour: models and challenges," *Transport Reviews*, vol. 27, no. 1, pp. 65–84, 2007.
- [10] M. Treiber and A. Kesting, *Traffic Flow Dynamics*, Springer, Berlin, Germany, 2013.
- [11] M. Saifuzzaman and Z. Zheng, "Incorporating human-factors in car-following models: a review of recent developments and research needs," *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 379–403, 2014.
- [12] E. Brockfeld, R. D. Kühne, and P. Wagner, "Calibration and validation of microscopic traffic flow models," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1876, pp. 62–70, 2004.
- [13] P. Ranjitkar, T. Nakatsuji, and M. Asano, "Performance evaluation of microscopic traffic flow models with test track data," *Transportation Research Record*, no. 1876, pp. 90–100, 2004.
- [14] E. Brockfeld, R. D. Kühne, and P. Wagner, "Calibration and validation of microscopic models of traffic flow," *Transportation Research Record*, vol. 1934, no. 1, pp. 179–187, 2005.
- [15] S. Panwai and H. Dia, "Comparative evaluation of microscopic car-following behavior," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 3, pp. 314–325, 2005.
- [16] S. Ossen and S. P. Hoogendoorn, "Heterogeneity in car-following behavior: theory and empirics," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 2, pp. 182–195, 2011.
- [17] I. Soria, L. Elefteriadou, and A. Kondyli, "Assessment of car-following models by driver type and under different traffic, weather conditions using data from an instrumented vehicle," *Simulation Modelling Practice and Theory*, vol. 40, pp. 208–220, 2014.
- [18] S. Bexelius, "An extended model for car-following," *Transportation Research*, vol. 2, no. 1, pp. 13–21, 1968.
- [19] H. Lenz, C. K. Wagner, and R. Sollacher, "Multi-anticipative car-following model," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 7, no. 2, pp. 331–335, 1999.
- [20] N. Eissfeldt and P. Wagner, "Effects of anticipatory driving in a traffic flow model," *The European Physical Journal B: Condensed Matter and Complex Systems*, vol. 33, no. 1, pp. 121–129, 2003.
- [21] S. P. Hoogendoorn and S. Ossen, "Empirical analysis of two-leader car-following behavior," *European Journal of Transport and Infrastructure Research*, vol. 6, no. 3, pp. 229–246, 2006.
- [22] M. Treiber, A. Kesting, and D. Helbing, "Delays, inaccuracies and anticipation in microscopic traffic models," *Physica A*, vol. 360, no. 1, pp. 71–88, 2006.
- [23] S. J. L. Ossen, *Longitudinal Driving Behavior: Theory and Empirics*, TRAIL Research School, Delft, The Netherlands, 2008.
- [24] S. P. Hoogendoorn, S. Ossen, and M. Schreuder, "Empirics of multianticipative car-following behavior," *Transportation Research Record*, vol. 1965, no. 1, pp. 112–120, 2006.
- [25] N. Farhi, H. Haj-Salem, and J.-P. Lebacque, "Multianticipative piecewise-linear car-following model," *Transportation Research Record*, vol. 2315, no. 1, pp. 100–109, 2012.
- [26] X. Zhang, "Empirical analysis of a generalized linear multianticipative car-following model in congested traffic conditions," *Journal of Transportation Engineering*, vol. 140, no. 6, Article ID 04014018, 2014.
- [27] X. Ma, "A neural-fuzzy framework for modeling car-following behavior," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, vol. 2, pp. 1178–1183, October 2006.
- [28] S. Panwai and H. Dia, "Neural agent car-following models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 60–70, 2007.
- [29] A. Khodayari, A. Ghaffari, R. Kazemi, and R. Braunstingl, "A modified car-following model based on a neural network model of the human driver effects," *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, vol. 42, no. 6, pp. 1440–1449, 2012.
- [30] L. Chong, M. M. Abbas, A. M. Flintsch, and B. Higgs, "A rule-based neural network approach to model driver naturalistic behavior in traffic," *Transportation Research Part C: Emerging Technologies*, vol. 32, pp. 207–223, 2013.
- [31] J. Zheng, K. Suzuki, and M. Fujita, "Car-following behavior with instantaneous driver-vehicle reaction delay: a neural-network-based methodology," *Transportation Research Part C: Emerging Technologies*, vol. 36, pp. 339–351, 2013.
- [32] D. Wei and H. Liu, "Analysis of asymmetric driving behavior using a self-learning approach," *Transportation Research Part B: Methodological*, vol. 47, pp. 1–14, 2013.
- [33] NGSIM Interstate 80 freeway dataset, 2014, <http://www.fhwa.dot.gov/publications/research/operations/06137/>.
- [34] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [35] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 4, pp. 276–281, 2004.
- [36] R. L. Cheu, J. Xu, A. G. H. Kek, W. P. Lim, and W. L. Chen, "Forecasting shared-use vehicle trips with neural networks and support vector machines," *Transportation Research Record*, vol. 1968, no. 1, pp. 40–46, 2006.
- [37] Y. Zhang and Y. Xie, "Forecasting of short-term freeway volume with v-support vector machines," *Transportation Research Record*, vol. 2024, no. 1, pp. 92–99, 2007.
- [38] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: where we are and where we're going," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, 2014.
- [39] H. Drucker, C. J. C. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Advances in Neural Information Processing Systems*, vol. 9, pp. 155–161, 1997.

- [40] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [41] I. IGuyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
- [42] B. Lu and S. S. Washburn, "Vehicle trajectory analysis system for estimating HCM compatible performance measures," in *Proceedings of the 93rd Annual Meeting on Transportation Research Board*, Washington, DC, USA, January 2014.
- [43] T. Toledo, H. N. Koutsopoulos, and K. I. Ahmed, "Estimation of vehicle trajectories with locally weighted regression," *Transportation Research Record*, vol. 1999, no. 1, pp. 161–169, 2007.
- [44] C. Thiemann, M. Treiber, and A. Kesting, "Estimating acceleration and lane-changing dynamics from next generation simulation trajectory data," *Transportation Research Record*, vol. 2088, pp. 90–101, 2008.
- [45] V. Punzo, M. T. Borzacchiello, and B. Ciuffo, "On the assessment of vehicle trajectory data accuracy and application to the Next Generation Simulation (NGSIM) program data," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1243–1262, 2011.
- [46] X. Ma and I. Andréasson, "Estimation of driver reaction time from car-following data application in evaluation of general motor-type model," *Transportation Research Record: Journal of the Transportation Research Board*, no. 1965, pp. 130–141, 2006.
- [47] J. Ma, J. Theiler, and S. Perkins, "Accurate on-line support vector regression," *Neural Computation*, vol. 15, no. 11, pp. 2683–2703, 2003.
- [48] R. S. Pindyck and D. L. Rubinfeld, *Econometric Models and Economic Forecasts*, McGraw-Hill, Boston, Mass, USA, 4th edition, 2000.
- [49] X.-Q. Chen, W.-J. Xie, J. Shi, and Q.-X. Shi, "Perturbation and stability analysis of the multi-anticipative intelligent driver model," *International Journal of Modern Physics C*, vol. 21, no. 5, pp. 647–668, 2010.
- [50] A. Kesting and M. Treiber, "Calibrating car-following models by using trajectory data methodological study," *Transportation Research Record: Journal of the Transportation Research Board*, no. 2088, pp. 148–156, 2008.
- [51] L. Cao, H. P. Lee, C. K. Seng, and Q. Gu, "Saliency analysis of support vector machines for gene selection in tissue classification," *Neural Computing & Applications*, vol. 11, no. 3-4, pp. 244–249, 2003.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

