

Research Article

Examining the Impact of Different Periodic Functions on Short-Term Freeway Travel Time Prediction Approaches

Xu Miao,¹ Bing Wu,¹ Yajie Zou ,¹ and Lingtao Wu²

¹The Key Laboratory of Road and Traffic Engineering, Ministry of Education, Tongji University, Shanghai 201804, China

²Texas A&M Transportation Institute, Texas A&M University System, College Station, TX 77843-3135, USA

Correspondence should be addressed to Yajie Zou; yajiezou@hotmail.com

Received 1 December 2019; Revised 8 July 2020; Accepted 11 July 2020; Published 1 August 2020

Academic Editor: Zhigang Xu

Copyright © 2020 Xu Miao 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.

Freeway travel time prediction is a key technology of Intelligent Transportation Systems (ITS). Many scholars have found that periodic function plays a positive role in improving the prediction accuracy of travel time prediction models. However, very few studies have comprehensively evaluated the impacts of different periodic functions on statistical and machine learning models. In this paper, our primary objective is to evaluate the performance of the six commonly used multistep ahead travel time prediction models (three statistical models and three machine learning models). In addition, we compared the impacts of three periodic functions on multistep ahead travel time prediction for different temporal scales (5-minute, 10-minute, and 15-minute). The results indicate that the periodic functions can improve the prediction performance of machine learning models for more than 60 minutes ahead prediction and improve the over 30 minutes ahead prediction accuracy for statistical models. Three periodic functions show a slight difference in improving the prediction accuracy of the six prediction models. For the same prediction step, the effect of the periodic function is more obvious at a higher level of aggregation.

1. Introduction

Travel time can effectively measure roadway traffic conditions [1]. Thus, accurate prediction of freeway travel time is important for traffic management agencies to provide better traffic guidance. However, it is challenging for researchers to predict travel time accurately due to the complex changes in traffic states [2]. A large number of algorithms have been proposed to improve the prediction accuracy of travel time. The existing short-term traffic forecasting algorithms were reviewed by Vlahogianni et al. [3] and Vlahogianni et al. [1]. Existing short-term traffic forecasting algorithms can be categorized into two major strands: statistical models and machine learning models. Linear regression analysis method [4, 5], time series method [6–8], and space time prediction methods [9, 10] are statistical models. Kalman filtering model [2, 11, 12], support vector regression model [13–15], and neural network model [16–20] are machine learning models. A series of combination models [21–25] are proposed in recent years.

Some researchers compared the performance of statistical models and machine learning models. For example, Stathopoulos et al. [26] found that fuzzy neural network outperformed Autoregressive Integrated Moving Average Model (ARIMA) in prediction performance. Vlahogianni [27] suggested that the advanced Neural Network (NN) structure can perform better than the ARIMA model. Jiang et al. [28] examined the prediction performance of different models under multiple steps ahead, and their results indicated that the machine learning models are superior to the two statistical models (i.e., vector autoregressive models and ARIMA).

Traffic data usually exhibit periodic characteristics during weekdays. Thus, considering the periodicity of data can improve the prediction performance. Up to date, three different approaches have been proposed to capture the periodic characteristics. Zou et al. [29] found that a synthetic prediction model consisting of statistical models and trigonometric polynomial function (TPF) can achieve higher prediction accuracy when the forecasting horizon is greater

than half hour with 5 minutes as the aggregation level. Tang et al. [30] applied a double exponential smoothing method (DES) to describe the weekly similarities of traffic data. In the course of the study, Chen et al. [31] utilized the prediction model in accordance with the original traffic flow series compared with the intraday trend removed the by simple average (SA) approach. It is found that the accuracy of the prediction could be considerably improved by using the residual time series.

Regarding the prediction interval (steps), some existing studies have investigated the impact of data resolution on model prediction performance, but there are no definitive results. For example, Park et al. [32] considered the aggregation level from 2 minutes to 60 minutes of the ARIMA model based on travel time data. They concluded that forecasting route travel time required higher concentration levels than link travel time prediction. Vlahogianni et al. [33] found that time clustering may distort critical traffic flow information, and we need further research to determine the optimal concentration level. Some studies found that higher data resolution usually shows larger noise [34, 35].

Based on the previous studies, some studies have compared statistical models and machine learning models, and some scholars have proposed the improvement of periodic functions on travel time prediction. However, few studies have comprehensively evaluated the effects of different periodic functions on the two types of models under different prediction steps. Thus, this study focuses on multistep ahead travel time prediction by considering different periodic functions. The periodic characteristics of the travel time are captured by SA, TPF, and DES models. The residual part is modeled by the statistical models (ARIMA, space time (ST) model, vector autoregressive (VAR) model) and machine learning models (support vector machine (SVM), back propagation neural network (BPNN), multilinear regression (MLR)). In total, 18 hybrid prediction models were established and compared. In addition, the performance of prediction models was evaluated under different scenarios: multistep ahead prediction (1, 3, 6, and 12 steps ahead predictions) with different aggregation levels (5-minute, 10-minute, and 15-minute).

The remainder of the paper is organized as follows. In Section 2, we introduce the travel time data in the study. We describe the data collection site and analyze the temporal and spatial correlation as well as the diurnal pattern observed in the data. In Section 3, we introduce periodic functions and two main methodologies used in this study: statistical models and machine learning approaches. We also discuss the evaluation measures and determine the appropriate training periods. In Section 4, we evaluate the prediction performance of the six models and compare the impacts of different periodic functions on prediction models under different scenarios. In Section 5, we provide the conclusions and some future works.

2. Travel Time Data

This study analyzed the travel time data of US-290 between IH-610 and FM-1960 in Houston, Texas. The total length is

approximately 12 miles. The segment is divided into five links by six automatic vehicle identification (AVI) readers (Figure 1). Vehicles with toll tags passed through the AVI readers will be recorded with their ID and timestamps. Travel time of the link enclosed by this pair of AVI readers is the difference in the timestamps. The length of link A to link E is 0.8, 2.6, 3.0, 1.5, and 4.1 miles, respectively. The data collection duration is from January 2008 to August 2008, a total of 174 days. The travel times were initially collected once every 30 seconds, 24 hours per day. We calculated the arithmetic mean of travel time and aggregated the travel time into 5-minute, 10-minute, and 15-minute intervals for each link. The missing data for the five links are all less than 1%, and historical averaged based data imputation method have been implemented to ensure the selected travel time data are appropriate for model validation and evaluation in this study. This study only focuses on the weekday (Monday-Friday) travel time prediction.

2.1. Temporal and Spatial Correlation of the Travel Time. We calculate the historical average travel time per mile of the five links (Monday to Friday, January to August 2008) (Figure 2). It can be found that the peak time of traffic occurs in the afternoons of all these five links, and there are mainly three types of travel time patterns. For link A, travel time increases after 12:00, peaks at about 16:30, and finishes later than 20:00. For links B and C, traffic congestion starts at 12:00, peaks around 17:35, and returns to usual after 20:00. For links D and E, traffic congestion often occurs before 16:00, peaks around 17:50, and dissipates after 20:00. In our study, link D is chosen as the target link.

Changes in traffic flow have certain temporal and spatial characteristics. Autocorrelation and cross-correlation functions were calculated to examine the temporal and spatial correlation. The equation adopted here follows that of Zou et al. [23], as shown in equations (1)–(3).

$$ar_{k_a} = \frac{ac_{k_a}}{ac_0}, \quad (1)$$

$$ac_{k_a} = \frac{1}{N_a} \sum_{t=1}^{N_a-k_a} (x_t - \bar{x})(x_{t+k_a} - \bar{x}),$$

where ar_{k_a} is the sample autocorrelation function; k_a is the time lag; $ac_0 = ac_{k_a=0}$; N_a is the number of observations; x_t is the sample observation; and \bar{x} is the sample means of the series.

In this case, the cross-correlation function measures the temporal and spatial correlation between the travel time data pairs recorded on two selected links. For travel time data pairs $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, an estimate of the cross-covariance function $c_{xy}(k_c)$ is

$$c_{xy}(k_c) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-k_c} (x_t - \bar{x})(y_{t+k_c} - \bar{y}), & k_c = 0, 1, 2, \dots, \\ \frac{1}{n} \sum_{t=1}^{n+k_c} (y_t - \bar{y})(x_{t-k_c} - \bar{x}), & k_c = 0, -1, -2, \dots, \end{cases} \quad (2)$$

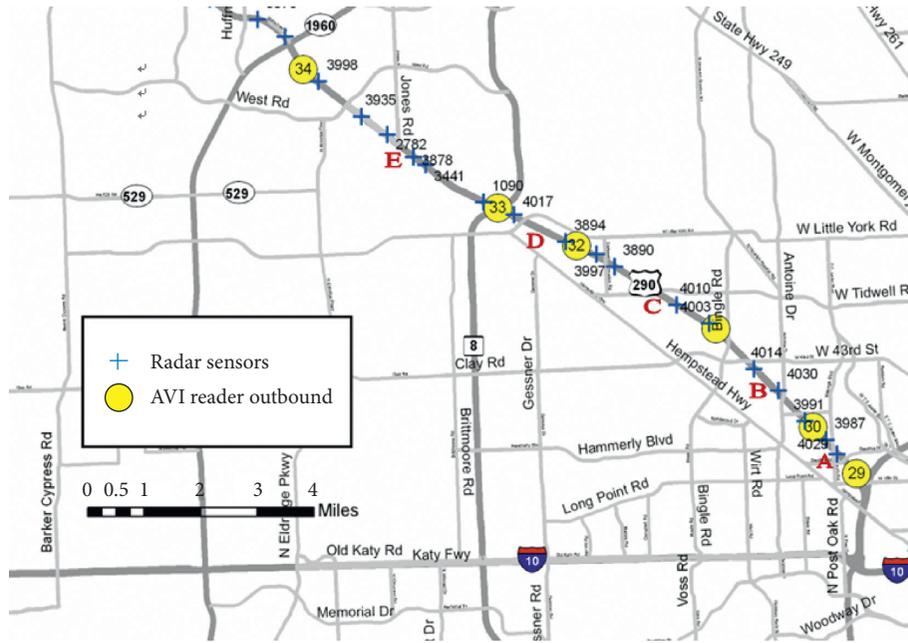


FIGURE 1: Study links along the US-290 [23].

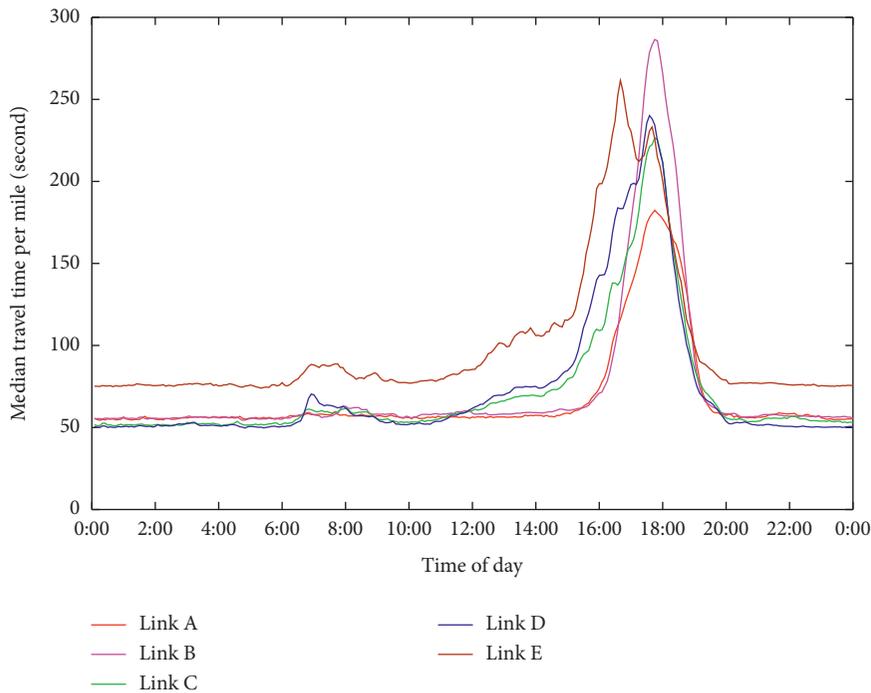


FIGURE 2: Historical median travel times on the five links.

where \bar{x} and \bar{y} are the sample means of the x_t series and y_t series, respectively. n is the number of travel time data pairs, and an estimate of the lag k_c cross-correlation function is

$$cr_{xy}(k_c) = \frac{c_{xy}(k_c)}{s_x s_y}, \quad k_c = 0, \pm 1, \pm 2, \dots, \quad (3)$$

where $s_x = \sqrt{c_{xx}(0)}$ and $s_y = \sqrt{c_{yy}(0)}$, and $cr_{xy}(k_c)$ is the sample cross-correlation function.

We found that the autocorrelation function of travel time shows a downward trend with time lag (Figure 3). Cross-correlation functions between link D and links A, B, C, and E peak at the lag of -9 and -4 , 0 and 0 , respectively (Figure 4). As can be seen from previous analysis (Figure 2), the peak times of five links are not same. The peak of links A and B occurs earlier than link D, so the cross-correlation functions between links A and B and link D peak reach the peak at lags of -9 and -4 . The traffic state of links A and B

changed 45 minutes and 20 minutes earlier than link D. Links C and E are directly connected to link D, the traffic congestion state and the peak time are more similar, the cross-correlation functions between links C, E and link D peak at lags of 0. Furthermore, the maximum cross-correlation values between link D and its adjacent links A, B, C, and E are 0.547, 0.720, 0.822, and 0.904.

2.2. Periodic Pattern of Travel Time. Previous research showed that travel time exhibits periodic characteristics during the weekdays. Similar periodic characteristics were found by Kamarianakis et al. [36] in occupancy, speed, and flow data. Because the periodic trend may affect the travel time prediction, this study proposed the hybrid prediction models to accommodate the periodic trend components as well as the temporal and spatial correlation observed in the data. Specifically, periodic characteristics are modeled using TPF, SA, and DES methods.

3. Methodology

3.1. Periodic Functions

3.1.1. Simple Average Method. Simple average method is one of the commonly used methods to describe the periodic characteristics [31]. During the study, the researchers set the hypothesis that the sampling travel time data of M consecutive working days could be written as a series of one-dimensional vectors Y_d , as shown in equation (4). The intraday trend \bar{Y} is calculated by simple average method as equation (5).

$$Y_1 = [y_1(1), y_1(2), \dots, y_1(m)], \dots \quad (4)$$

$$Y_M = [y_M(1), y_M(2), \dots, y_M(m)],$$

$$\bar{Y} = [\bar{y}(1), \dots, \bar{y}(m)] = \left[\frac{1}{M} \sum_{d=1}^{d=M} y_d(1), \dots, \frac{1}{M} \sum_{d=1}^{d=M} y_d(m) \right], \quad (5)$$

where $y_d(m)$ stands for travel time data collected at time m on day d . M indicates the number of sampled days, m is sampling data points per day. In this study, $M = 30$, $m = 288$.

3.1.2. Trigonometric Polynomial Function. Trigonometric polynomial adopts the sinusoids and cosinusoids to describe the periodic pattern. Equation (6) was used to calculate the average daily travel time at each station. Trigonometric polynomial function is represented in the following equation:

$$S_t = a_0 + a_1 \sin\left(\frac{2\pi t}{m}\right) + a_2 \cos\left(\frac{2\pi t}{m}\right) + \dots \quad (6)$$

$$+ a_{2n_r-1} \sin\left(\frac{2n_r \pi t}{m}\right) + a_{2n_r} \cos\left(\frac{2n_r \pi t}{m}\right),$$

where S_t is the estimated periodic component at time t ; m indicates the number of samples per day; n_r is the number of

trigonometric polynomials; and a_0, \dots, a_{2n_r} are the coefficients.

Regarding the selection of optimal number of trigonometric series functions, Zou et al. [29] claimed that the number of trigonometric polynomials might have an impact on the prediction accuracy of the hybrid model. At the same time, they found that 15 or more trigonometric polynomials should be included in the periodical component. Therefore, in this study, the researchers set the value of n_r in equations (6)–(15).

3.1.3. Double Exponential Smoothing. Double exponential smoothing is one widely used method for both smoothing and forecasting time series. This approach builds the prediction in accordance of the levels mean M_t and the trend T_t . The model can be expressed as

$$M_t = \alpha Y_t + (1 - \alpha)(M_{t-1} + T_{t-1}),$$

$$T_t = \beta(M_t - M_{t-1}) + (1 - \beta)T_{t-1},$$

$$Y_{t+1} = M_t + T_t, \quad (7)$$

$$M_1 = Y_1,$$

$$T_1 = \frac{((Y_2 - Y_1) + (Y_3 - Y_2) + (Y_4 - Y_3))}{3},$$

where Y_t is the observed travel time at time t ; M_t stands for the estimate of level of series at time t ; T_t indicates the estimate of slope of series at time t ; α and β are smoothing parameters, the two parameters can be estimated using the Levenberg-Marquardt algorithm.

3.2. Travel Time Prediction Models

3.2.1. Statistical Models

(1) Autoregressive Integrated Moving Average (ARIMA) Models. The ARIMA model transforms nonstationary time series into stationary time series after di differences, and then stationary sequence can be predicted by the ARMA model. From the view of mathematics, the demonstration of an ARMA (p, q) procedure is as

$$y(t) = \varepsilon(t) + \sum_{i=1}^{i=p} \varphi_i y(t-i) + \sum_{j=1}^{j=q} \theta_j \varepsilon(t-j), \quad (8)$$

where $y(t)$ stands for the future travel time at time t ; φ_i and θ_i are the parameters of pattern; $\varepsilon(t)$ indicates white Gaussian noise with mean zero and variance σ_w^2 ; p is the number of autoregressive terms; q is the amount of lagged forecast errors. Let

$$B^i y(t) = y(t-i). \quad (9)$$

ARIMA model is as equation (10):

$$\left(1 - \sum_{i=1}^{i=p} \varphi_i B^i\right) (1 - B)^{di} y(t) = \left(1 + \sum_{j=1}^{j=q} \theta_j B^j\right) \varepsilon(t), \quad (10)$$

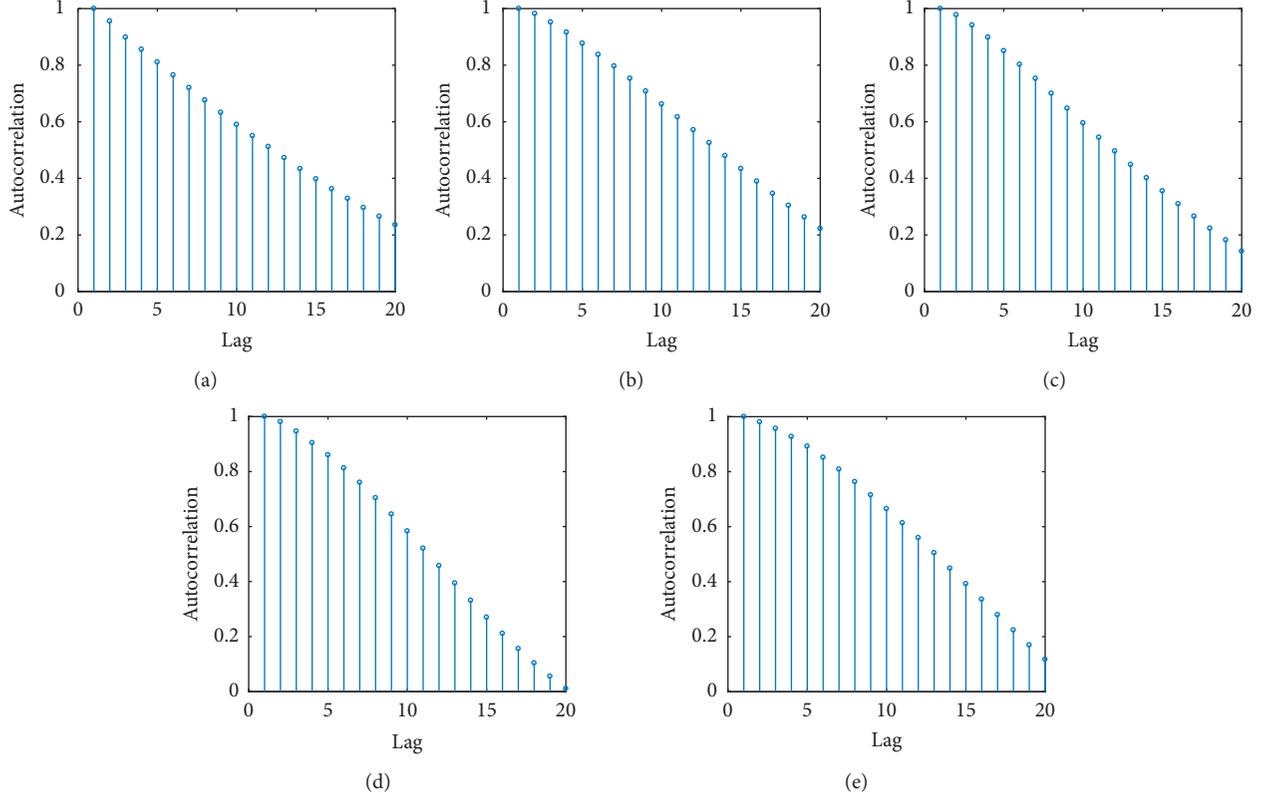


FIGURE 3: Autocorrelation functions of 5-minute travel time on links A, B, C, D, and E. (a) Link A. (b) Link B. (c) Link C. (d) Link D. (e) Link E.

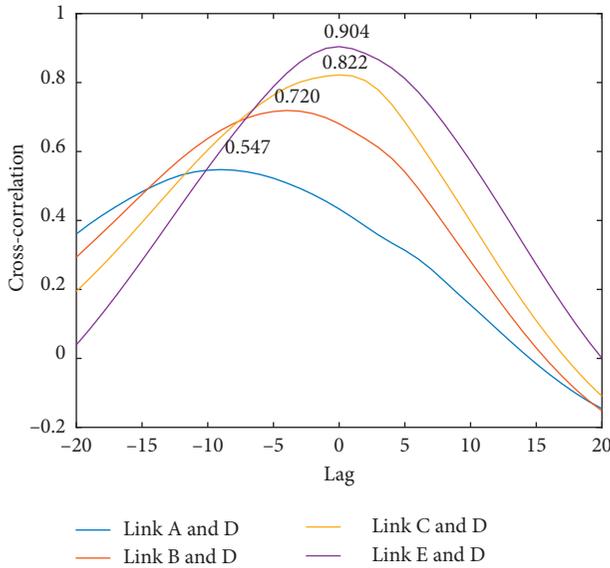


FIGURE 4: Cross-correlation functions of 5-minute travel time on link D.

where di is a nonnegative integer, which stands for the number of nonseasonal differences. If $di = 0$, the ARMA model could be obtained. When predicting each future travel time value, the best order of the ARIMA model is decided by Akaike information criterion (AIC).

(2) *Space Time Model*. As a probabilistic modeling method, ST model can provide point prediction and corresponding prediction intervals. The normal distribution is used to describe travel time in this study. The point prediction D_{t+p} is D_{t+p} given by

$$Z_\tau = u_{t+pr} + \sigma_{t+pr} \times \Phi^{-1}[\tau], \quad (11)$$

where $pr = 1, 2, 3, \dots, 12$; u_{t+pr} and σ_{t+pr} are the location parameter and scale parameter of $N(u_{t+pr}, \sigma_{t+pr}^2)$; Φ is the cumulative density function of a standard normal distribution. The u_{t+pr} is modeled through a linear combination of current and previous values of all travel time series on all links. When choosing the predictive variables, different combinations of predictive variables need to be considered. Therefore, researchers begin with the most complex models and gradually subtract predictive variables until no further improvement was obtained. For instance, if $pr = 1, 9$ variables were selected,

$$u_{t+1} = \alpha_0 + \alpha_1 A_t + \alpha_2 B_t + \alpha_3 B_{t-1} + \alpha_4 C_t + \alpha_5 C_{t-1} + \alpha_6 D_t + \alpha_7 D_{t-1} + \alpha_8 E_t + \alpha_9 E_{t-1}, \quad (12)$$

where A_t, B_t, C_t, D_t, E_t are the travel time at links A, B, C, D, and E at time t and $\alpha_0, \alpha_1, \dots, \alpha_9$ are model coefficients.

To build a model for the predictive spread, σ_{t+p} , the ST model allows for conditional heteroscedasticity by modeling σ_{t+p} as a linear function of the volatility value v_t ,

$$\sigma_{t+pr} = b_0 + b_1 v_t. \quad (13)$$

The coefficients b_0 and b_1 are nonnegative, and their volatility values could be modeled as

$$v_t = \left(\frac{1}{10} \sum_{i=0}^t ((A_{t-i} - A_{t-i-1})^2 + (B_{t-i} + B_{t-i-1})^2 + (C_{t-i} + C_{t-i-1})^2 + (D_{t-i} + D_{t-i-1})^2 + (E_{t-i} + E_{t-i-1})^2) \right)^{1/2}. \quad (14)$$

(3) *Vector Autoregressive Models*. VAR model is regarded as one of the most widespread methods which utilize statistical methods in time series prediction. The model can include many factors consisting of the impact of upstream and downstream links on predicting future travel time. During the process of this research, a 5-equation VAR model is utilized, and it can be expressed as follows:

$$X_{t+1} = c_0 + c_1 X_t + c_2 X_{t-1} + \dots + c_k X_{t-k} + f_{t+1}, \quad (15)$$

where $X_{t+1} = (A_{t+1}, B_{t+1}, C_{t+1}, D_{t+1}, E_{t+1})^T$; $c_0 = 5 \times 1$ constant term; $c_1, \dots, c_k = 5 \times 5$ coefficient matrices; and f_{t+1} = the corresponding 5×1 independently and identically distributed random vector, $E(f_{t+1}) = 0$.

The stability of the VAR model could be guaranteed through the characteristic polynomial

$$\det(I_5 - c_1 z - \dots - c_k z^k) \neq 0, \quad \text{for } |z| \leq 1, \quad (16)$$

where I_5 stands for a 5×5 identity matrix. It is a necessary and sufficient condition that all characteristic roots are located outside the unit circle for stability.

3.2.2. Machine Learning Models

(1) *Support Vector Machine Model*. The SVM approach is a method which could be used to map the sample space into high or even infinite dimension feature space (Hilbert space) by nonlinear mapping to construct linear regression in a new space. Given a set of data points $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ for regression, N is the number of training samples. Normally, the objective of SVM is to find a function

$$g(x) = w \cdot \Phi(x) + b, \quad (17)$$

where $\Phi(\cdot)$ = the kernel function that maps input x into the feature space g ; w is the weighting vector; b is a constant bias.

A λ -insensitive loss function is assumed as

$$L_\lambda(x, y, g) = |y - g(x)|_\lambda = \max(0, |y - g(x)| - \lambda). \quad (18)$$

Then, it could be estimated that w and b by working out this optimization problem:

$$R(f) = C \sum_{i=1}^{i=N} L_\lambda(x_i, y_i, g) + \frac{1}{2} \|w\|^2, \quad (19)$$

where λ is the maximum deviation permitted; $L_\lambda(x, y, g)$ is loss function, C indicates the related penalty for stating

deviation within the training process that assesses the tradeoff between the empirical risk and the smoothness of the model. The relaxation variables β_i and β_i^* are used to indicate the optimization objective into the optimization issue stated as

$$\begin{aligned} \min \quad & C \sum_{i=1}^{i=N} (\beta_i + \beta_i^*) + \frac{1}{2} \|w\|^2 \\ & y_i - w \cdot \Phi(x) - b \leq \lambda + \beta_i \\ \text{s.t.} \quad & w \cdot \Phi(x) + b - y_i \leq \lambda + \beta_i^* \\ & \beta_i, \beta_i^* \geq 0. \end{aligned} \quad (20)$$

That issue mentioned above is worked out by utilizing the Lagrange equation. The regression function is demonstrated as

$$f(x) = \sum_{x_i \in SV} (\alpha_i - \alpha_i^*) E(s_i, s_j) + b, \quad (21)$$

$$E(s_i, s_j) = \exp(-\gamma \|s_i, s_j\|^2),$$

where $E(s_i, s_j)$ is the kernel function. α_i and α_i^* are the solutions to dual problem. In our study, the grid analysis and cross validation are used to optimize the parameters C and γ . The cross-validation method divides data into three groups, among which one subset is the validation group and the other two subsets are used as the training set; 3 models are obtained. Grid analysis is a method of programming enumeration to compare the performance of models with different parameters C and γ . In this paper, all combinations of $\log 2C$ and $\log 2\gamma$ parameters between -5 and 5 were traversed. The parameter combination with minimum mean square error was selected.

(2) *Back Propagation Neural Network Model*. In short, the BPNN model is a multilayer feed forward neural network which consists of many parallel nonlinear computing elements. As we all know, initialization network is composed of input layer, hidden layer, and output layer. Within the neural network, the weights between the most important parameter connection layers can be calculated by error back propagation algorithm. When a neural network model acquires the mapping relationship between input and output variables through continuous learning, it can predict the output according to the given input variables.

First, equation (22) can be used to calculate the value of the predicted hidden layer:

$$H_j = S \left(\sum_{i=1}^{i=\text{num}} v_{ij} x_j - e_j \right), \quad j = 1, 2, \dots, h, \quad (22)$$

where H_j stands for the production of hidden layer and S is the incentive function of neurons, h stands for the neuron number of hidden layers, num refers to the neuron number of the input-layers, v_{ij} stands for the weight element between input-layer and hidden-layer, e_j stands for the bias value of hidden layer.

Second, predicting value of the output layer could be calculated through

$$Q = \sum_{j=1}^{j=h} H_j v_{j\zeta} - q_\zeta, \quad \zeta = 1, 2, \dots, k_1, \quad (23)$$

where Q stands for the actual output of output layer, q_ζ refers to the bias value of output layer, and k_1 stands for the neuron number of the output layer. In our study, the empirical formula combined with the trial and error method was used to determine the number of nodes in the hidden layer. 4 nodes with the best performance were selected finally.

(3) *Multilinear Regression Model*. Compared with the above two supervised algorithms, the construction of multiple linear regressions is simpler and belongs to regression learning category. In MLR, the prediction values can be calculated by the following equation:

$$y(t) = r_0 + \sum_{j=1}^{j=lr} r_j y(t-j), \quad (24)$$

where $y(t)$ represents the prediction value at time t . The independent variable $y(t-j)$ means the travel time data at the previous $t-j$ period, lr is the number of historical travel times considered in MLR model, and r_0, \dots, r_j are the regression parameters which can be optimized by training samples. lr is chosen on basis of an analysis of the travel data from January to April 2008. Different numbers of lr are considered.

3.3. Hybrid Prediction Models. As mentioned in Section 2, freeway travel time has a daily periodic characteristic. Therefore, it can be assumed that the travel time has two parts. One of the two parts is the deterministic component; the other is the irregular component. In such a hypothesis, the hybrid prediction model can be used to describe or calculate the freeway travel time:

$$P_t = D_t + y_t^r, \quad (25)$$

where P_t is the travel time at time t at station D; D_t is the periodic component; and y_t^r represents the residual part after removing the periodic component.

Periodic component can be described by three kinds of functions (TPF, SA, and DES), and the residual part is modeled by six prediction models. We compare the impacts of different periodic functions on multistep ahead freeway travel time prediction models using travel time data with different aggregation levels.

3.4. Measures and Training Period. To evaluate the multistep prediction performance of all prediction models, three indicators, mean absolute error (MAE), mean absolute prediction error (MAPE) and root mean square error (RMSE) are considered comprehensively. The equations for calculating three indexes are as follows:

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \\ \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| * .100\%, \\ \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, \end{aligned} \quad (26)$$

where n is the number of observations; y_i represents the actual travel time at time i on link D; and \hat{y}_i refers to the predicted travel time.

So far, there is no automatic way to calculate and evaluate the model training period. This study considered different training periods of 15, 20, 25, 30, 40, 50 and 60 days. For comparison, the travel time data in August (21 weekdays) were used as the test set. Figure 5 shows MAE, MAPE, and RMSE values of six travel time prediction models under different lengths of training periods. It is observed that the prediction performance of statistical models and MLR model changed slightly as the number of training period increases. The performance of SVM and BPNN has been greatly improved as training period increases when the training period was less than 30 days. If the training period was more than 30 days, the prediction accuracy of SVM and BPNN models changed slightly. Longer training period usually requires larger computational time for each model. For example, the computational time of the SVM model is 5 minutes when 10-day travel time data was used for model training, and the computational time can be as high as 68 minutes when 60 days was chosen as training period. The calculation time and prediction accuracy were considered comprehensively in our study, and a 30-day (July (23 days) and June (7 days)) training period is chosen for models.

4. Results and Discussion

In this part, the multistep ahead prediction performance of SVM, BPNN, MLR, ARIMA, ST, VAR under different aggregation levels (i.e., 5-minute, 10-minute, and 15-minute) are evaluated using the travel time data observed on link D. In addition, we explored the impacts of different periodic functions on statistical models and machine learning models under different aggregation levels for the input data. The testing period is 15:30 to 19:30 from 1 August to 31 August (21 weekdays).

4.1. The Performance of Six Models. The study provides the MAE, MAPE, and RMSE values of SVM, BPNN, MLR, ARIMA, ST, and VAR models for different forecasting horizons under different aggregation levels (5-minute, 10-minute, and 15-minute) for the input data (Tables 1–3). Tests on travel time data indicate the following findings. First, the prediction accuracy deteriorates as the forecasting step increases for all models. Second, the higher the data aggregate level, the higher the accuracy of short-term travel time

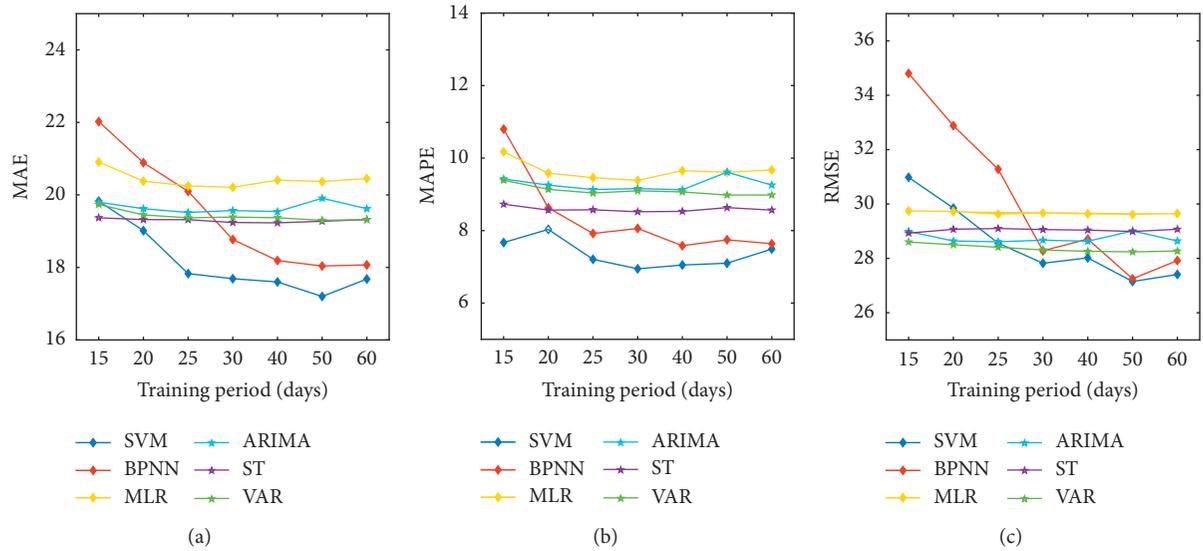


FIGURE 5: Comparison of training period for the different models.

TABLE 1: MAE, MAPE, and RMSE of six models for different minutes ahead predictions with 5-minute as aggregating level.

Time scale: 5 min		Minutes ahead predictions						
	Model	5 min	10 min	15 min	30 min	45 min	60 min	
MAE(s)	SVM	17.69	25.84	32.52	46.33	53.78	60.73	
	Machine learning models	BPNN	18.10	26.10	34.87	48.02	54.07	60.94
	MLR	20.21	34.01	46.32	74.53	95.71	111.05	
	ARIMA	19.80	31.13	40.56	57.02	65.98	77.50	
Statistical models	ST	19.37	29.86	38.40	51.22	63.41	73.79	
	VAR	19.39	30.74	39.31	56.48	68.82	78.16	
MAPE (%)	Model	5 min	10 min	15 min	30 min	45 min	60 min	
	SVM	6.95	10.18	12.69	19.90	25.14	29.73	
Machine learning models	BPNN	7.20	10.57	14.73	23.77	26.28	30.82	
	MLR	9.38	17.66	25.16	43.13	55.28	64.55	
	ARIMA	9.42	15.90	21.66	32.78	38.26	45.21	
Statistical models	ST	8.73	14.30	19.21	25.38	32.34	38.60	
	VAR	9.10	15.15	19.72	28.19	33.16	38.06	
RMSE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min	
	SVM	27.82	41.08	52.33	72.33	83.07	91.32	
Machine learning models	BPNN	28.38	40.93	53.60	73.15	81.75	89.47	
	MLR	29.67	46.96	61.95	92.99	118.52	134.65	
	ARIMA	28.99	44.47	56.80	77.89	92.55	108.58	
Statistical models	ST	28.93	43.57	55.60	73.85	94.27	108.50	
	VAR	27.98	43.62	55.18	78.74	98.34	105.69	

prediction results. For example, when we predict the 30-minute ahead travel time of link D, the prediction result is better with 15-minute data as the aggregation level than that with 10-minute and 5-minute data as the aggregation level. Third, for machine learning models, the prediction accuracy of MLR is lower than that of BPNN and SVM. For statistical models, the prediction accuracy of ARIMA is lower than that of ST and VAR. The possible reason is that the SVM, BPNN, ST, and VAR models use spatial and temporal information from neighboring links to predict the future travel time value at time $t + p$. While MLR and ARIMA models use the travel time data collected on the target link D only to predict travel time values at time $t + p$. Fourth, the prediction accuracy of

two machine models (SVM and BPNN) is better than that of statistical models. However, the prediction accuracy of MLR model is significantly lower than that of statistical models.

4.2. Impacts of Periodic Functions. To investigate whether the proposed periodic functions improve the performance of six prediction models, hybrid models and prediction models are considered to predict travel time values at link D for the same testing period (15:30 to 19:30 from 1 August to 31 August). The periodic function shows a consistent rule on improving the prediction model, whether it is based on MAE index, MAPE or RMSE index. The figure shows the

TABLE 2: MAE, MAPE, and RMSE of six models for different minutes ahead predictions with 10-minute as aggregating level.

Time scale: 10 min		Minutes ahead predictions					
MAE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min
Machine learning models	SVM	—	21.79	—	39.29	—	54.90
	BPNN	—	21.60	—	44.41	—	59.36
	MLR	—	30.32	—	71.39	—	109.57
Statistical models	ARIMA	—	29.32	—	62.79	—	86.27
	ST	—	26.88	—	51.55	—	71.44
	VAR	—	26.89	—	51.77	—	67.50
MAPE (%)	Model	5 min	10 min	15 min	30 min	45 min	60 min
Machine learning models	SVM	—	8.88	—	17.20	—	25.55
	BPNN	—	8.92	—	20.81	—	30.91
	MLR	—	15.51	—	41.05	—	64.00
Statistical models	ARIMA	—	14.93	—	36.55	—	50.69
	ST	—	12.44	—	26.19	—	37.60
	VAR	—	13.49	—	26.79	—	34.04
RMSE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min
Machine learning models	SVM	—	35.27	—	63.21	—	84.38
	BPNN	—	35.67	—	70.50	—	87.33
	MLR	—	41.98	—	89.53	—	132.75
Statistical models	ARIMA	—	41.49	—	83.54	—	115.66
	ST	—	40.09	—	72.82	—	104.54
	VAR	—	38.66	—	73.83	—	97.83

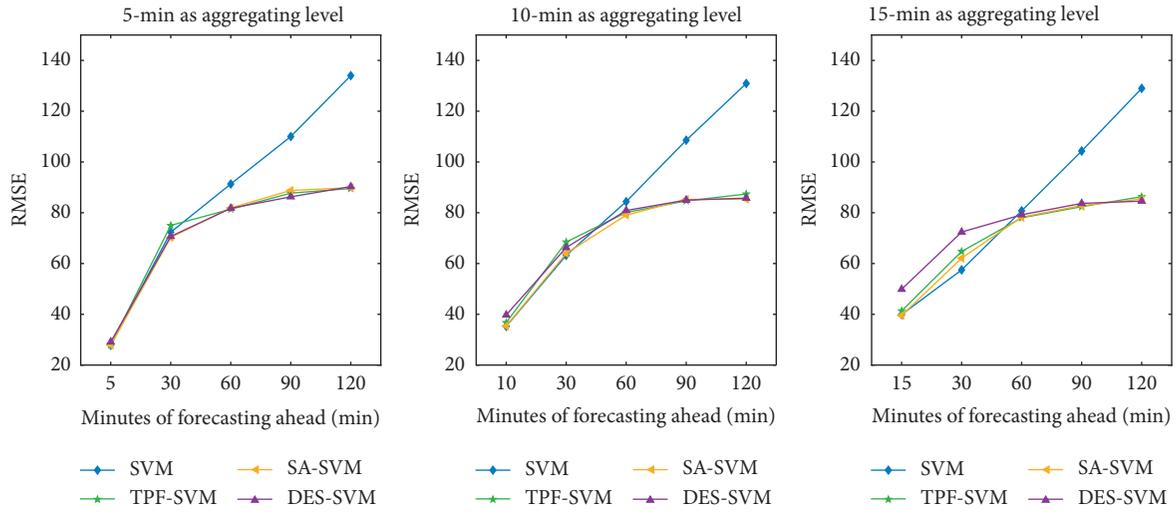
TABLE 3: MAE, MAPE, and RMSE of six models for different minutes ahead predictions with 15-minute as aggregating level.

Time scale: 15 min		Minutes ahead predictions					
MAE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min
Machine learning models	SVM	—	—	24.30	35.98	45.79	52.72
	BPNN	—	—	25.41	38.48	45.73	61.38
	MLR	—	—	37.14	68.42	88.75	106.79
Statistical models	ARIMA	—	—	32.64	56.16	65.82	79.20
	ST	—	—	30.47	49.01	60.88	68.28
	VAR	—	—	30.87	48.66	58.80	65.82
MAPE (%)	Model	5 min	10 min	15 min	30 min	45 min	60 min
Machine learning models	SVM	—	—	9.70	15.43	21.39	24.81
	BPNN	—	—	11.09	16.25	21.66	28.05
	MLR	—	—	20.99	39.86	51.47	62.47
Statistical models	ARIMA	—	—	18.19	33.01	37.99	48.02
	ST	—	—	15.55	24.54	31.12	35.59
	VAR	—	—	16.48	25.19	29.74	33.13
RMSE(s)	Model	5 min	10 min	15 min	30 min	45 min	60 min
Machine learning models	SVM	—	—	39.90	57.49	70.21	80.76
	BPNN	—	—	41.57	65.63	72.81	95.45
	MLR	—	—	49.56	85.06	112.00	130.03
Statistical models	ARIMA	—	—	45.50	75.91	90.63	106.43
	ST	—	—	45.06	71.97	91.30	100.50
	VAR	—	—	43.86	70.56	87.05	96.55

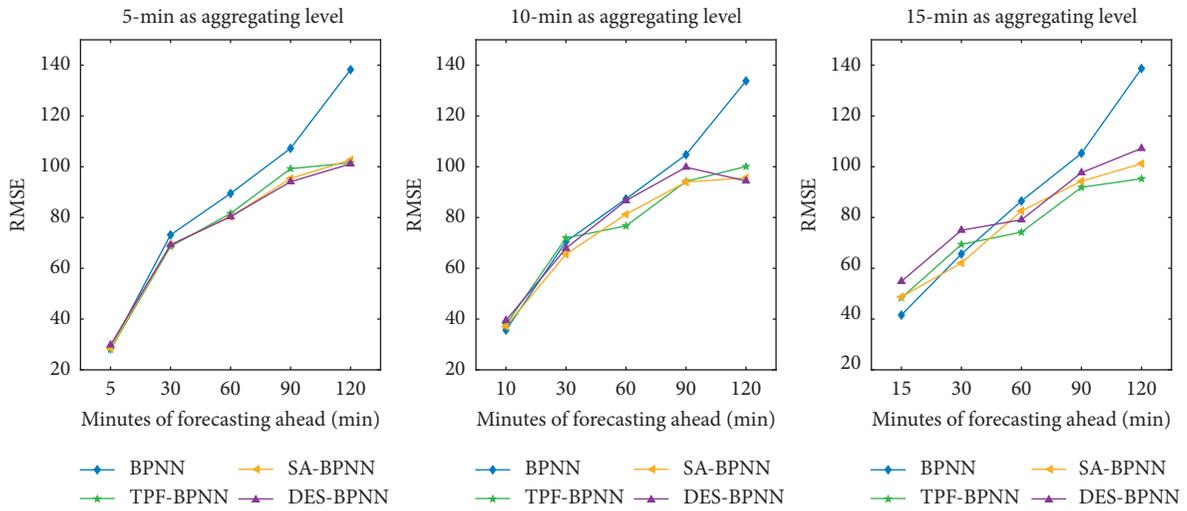
Bold values indicate the smallest MAE, MAPE, and RMSE values in machine learning models and statistical models, respectively.

RMSE results of the six models and 18 hybrid models for different forecasting horizons under different aggregation levels (5-minute, 10-minute, 15-minute) (Figure 6). Based on the observation of the results, several interesting conclusions can be drawn. First, period functions have similar impacts on SVM and BPNN models. The periodic functions have a definite improvement for more than 60 minutes ahead prediction under three data aggregation levels. Second, three periodic functions have improved the prediction performance of MLR model in multistep ahead

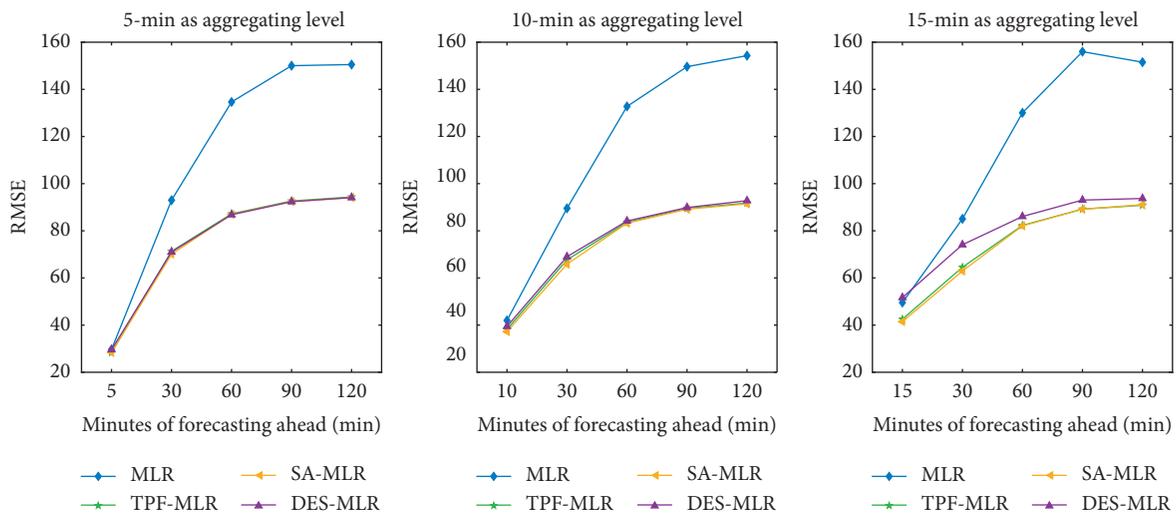
prediction for three data aggregation levels. Third, three periodic functions can improve the prediction accuracy of travel time over 30 minutes ahead for all statistical models. Fourth, with the increase of aggregation level, the difference of prediction results of the comprehensive prediction model considering periodic functions increases gradually. For example, when the aggregation level is 5-minute, the prediction results of the three SVM comprehensive models have little difference, while when the aggregation level is 15 minutes, the prediction results are significantly different.



(a)



(b)



(c)

FIGURE 6: Continued.

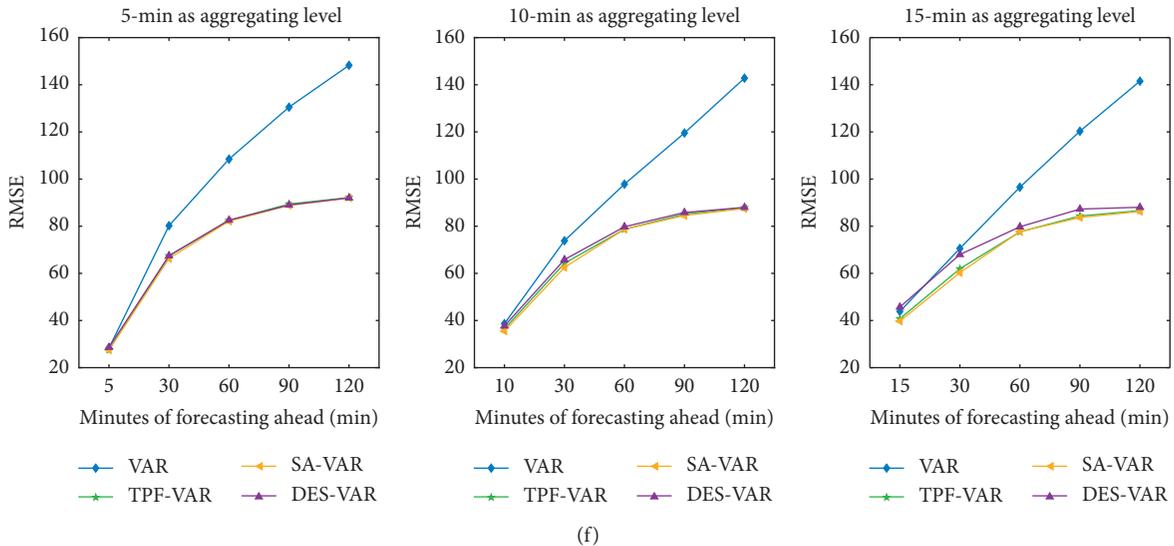
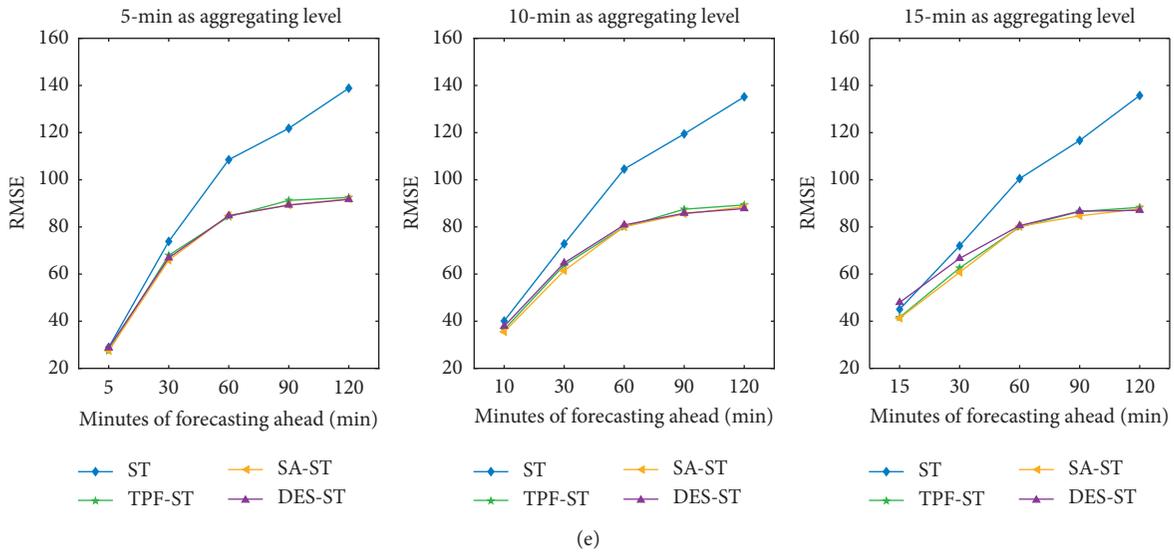
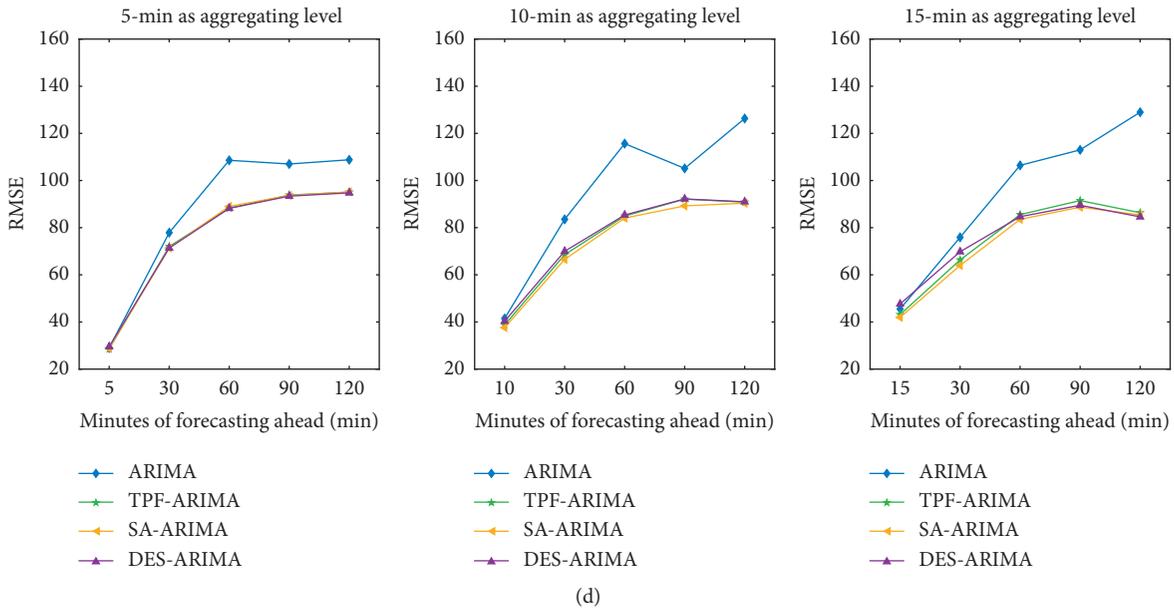
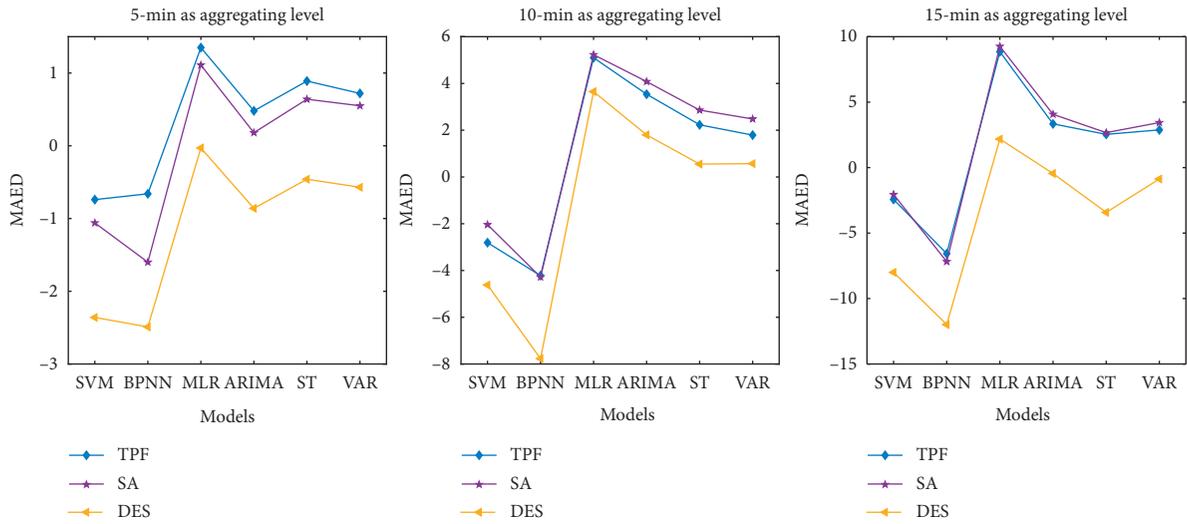
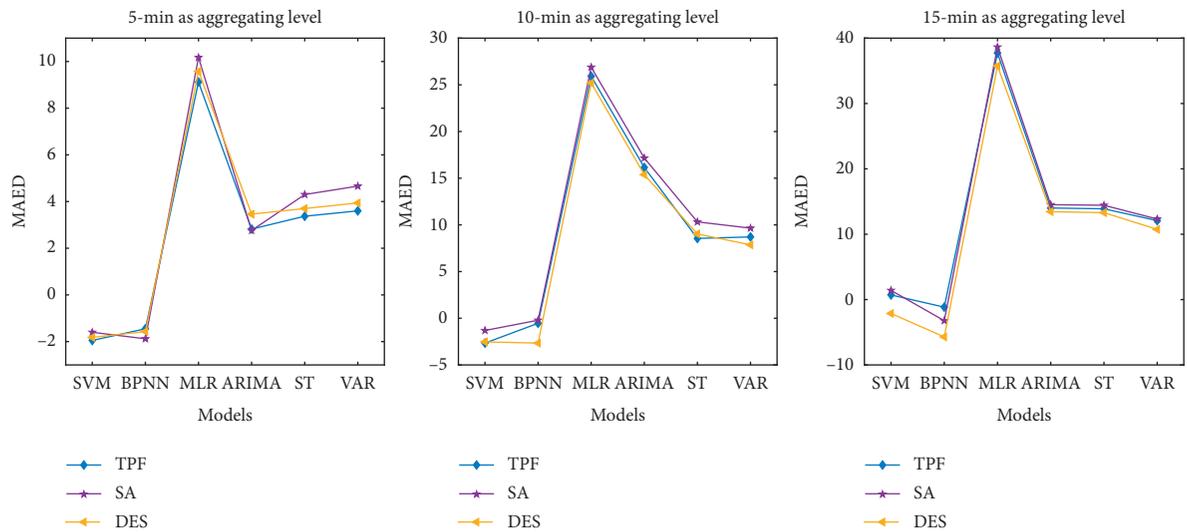


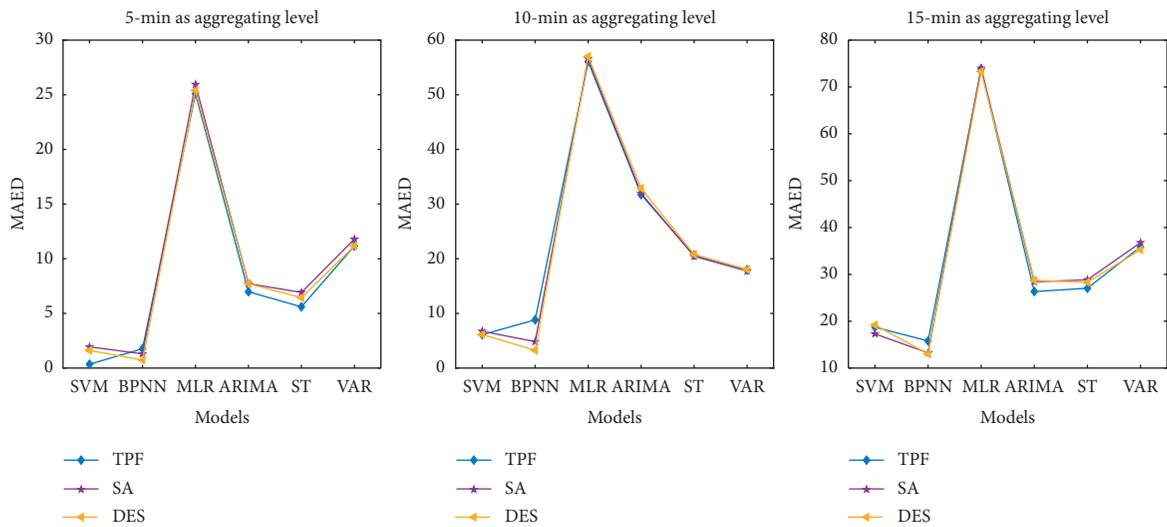
FIGURE 6: RMSE values of six models for different minutes ahead predictions. (a) RMSE values of SVM models for different minutes ahead predictions. (b) RMSE values of BPNN models for different minutes ahead predictions. (c) RMSE values of MLR models for different minutes ahead predictions. (d) RMSE values of ARIMA models for different minutes ahead predictions. (e) RMSE values of ST models for different minutes ahead predictions. (f) RMSE values of VAR models for different minutes ahead predictions.



(a)



(b)



(c)

FIGURE 7: Continued.

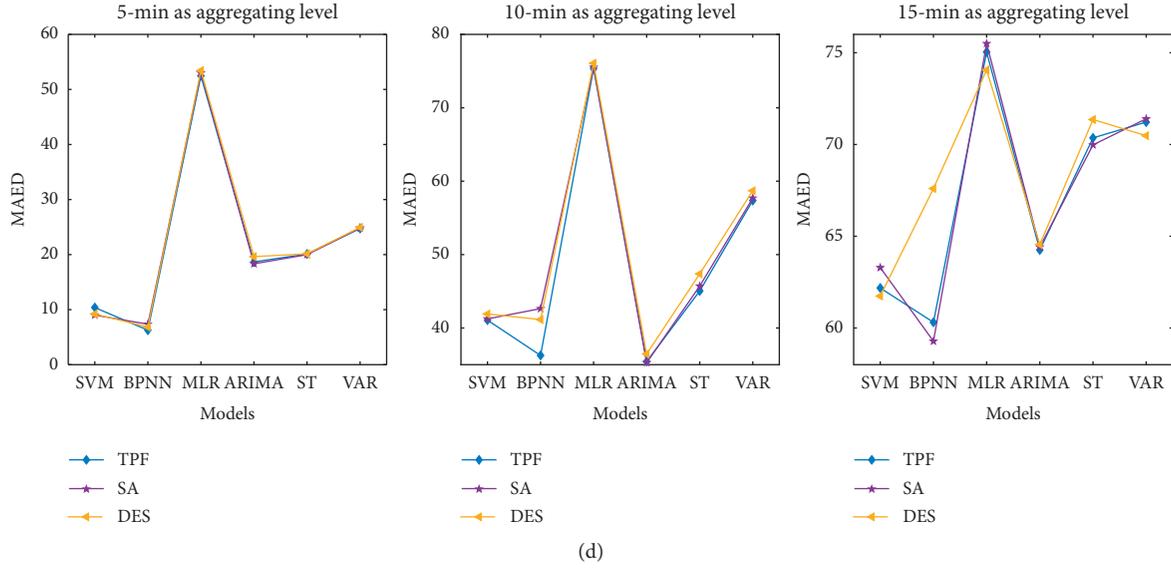


FIGURE 7: MAED for multistep ahead prediction with different aggregation levels (5-minute, 10-minute, and 15-minute). (a) MAED for 1-step ahead prediction. (b) MAED for 3-step ahead prediction. (c) MAED for 6-step ahead prediction. (d) MAED for 12-step ahead prediction.

TABLE 4: Aggregation level suggestions and periodic function suggestions.

Models	Minutes forecasting ahead (min)	Aggregating level suggestions (min)	Periodic function suggestions
SVM, BPNN	5	5	—
	10	10	—
	15	15	—
	30	15	—
	45	15	—
	60	15	TPF
	90	15	TPF
	120	15	TPF
ARIMA, ST, VAR, MLR	5	5	TPF
	10	10	SA
	15	15	SA
	30	15	SA
	45	15	SA
	60	15	SA
	90	15	SA
	120	15	SA

From the above analysis, we can conclude that the periodic functions obviously improve the prediction accuracy of the six prediction models for multistep ahead prediction. Then we analyze the impact degree of different periodic functions on prediction models based on mean absolute error difference (MAED). The equation of the MAED is as follows:

$$\text{MAED} = \frac{1}{n} \sum_{i=1}^n \left| \hat{y}_i - y_i \right| - \frac{1}{n} \sum_{i=1}^n \left| \hat{y}_{if} - y_i \right|, \quad (27)$$

where n is the number of observations; y_i represents the actual travel time at time i on link D; \hat{y}_i refers to the predicted travel time based on traditional prediction models, and \hat{y}_{if} refers to the predicted travel time based on models

considering periodic functions. If the MAED is greater than zero, this periodic function improves the prediction accuracy of the traditional prediction model; otherwise, it reduces the prediction accuracy of the traditional prediction model. The result shows the MAED values for 18 hybrid models from 1-step to 12-step ahead forecasting with 5-minute, 10-minute and 15-minute as aggregation levels (Figure 7). First, periodic functions can significantly improve the MLR model in multistep ahead prediction for three data aggregation levels. Second, for 1-step and 3-step ahead prediction, three periodic functions reduce the prediction accuracy of SVM and BPNN models. Third, for 1-step ahead prediction, both TPF and SA improve the performance of statistical models in multistep ahead prediction

for three data aggregation. When the aggregation level is 5-minute, TPF can obviously improve the prediction accuracy. While when the aggregation level increases to 10-minute, SA periodic function performs better. For 3-step ahead prediction, three periodic functions improve the forecasting results of the statistical models obviously, and three periodic functions have slight difference in improving prediction accuracy of the statistical models. Fourth, when 6-step and 12-step prediction ahead, three periodic functions improve the forecasting results of the six models obviously, and three periodic functions have slight difference in improving prediction accuracy of the prediction model. Fifth, for the same prediction step, the improvement of periodic function is more obvious with the increase of data aggregation level. For example, for 12-step ahead prediction, hybrid models perform better with 15-minute as aggregated level than that of with 5-minute as aggregated level.

In this section, we discuss the aggregation level and periodic function suggestions. According to the conclusion of Table 1 and Figure 7, the higher the data aggregate level, the higher the accuracy of short-term travel time prediction results. When prediction time was greater than 15 minutes, the highest accuracy could be obtained by using the aggregation level of 15 minutes. According to Figures 6 (a) and 6(b), the periodic functions cannot improve the prediction performance of SVM and BPNN models when prediction time was less than 60 minutes, it is recommended that no periodic function should be considered. When the prediction time is greater than 60 minutes, TPF periodic functions are recommended. As can be seen from Figure 7, both TPF and SA have improved the performance of statistical models and MLR model in multistep ahead prediction for three data aggregation and SA performed better. For different minutes ahead prediction, aggregation level and periodic function suggestions are shown in Table 4.

5. Conclusions

This paper evaluated the multistep ahead prediction performance of SVM, BPNN, MLR, ARIMA, ST, and VAR models using the freeway travel time data collected from vehicle identification readers along US-290 in Houston, Texas. The performances of the six prediction models under different aggregation levels (5-minute, 10-minute, and 15-minute) were compared. The impacts of different periodic functions on machine learning and statistical models under different aggregation levels (5-minute, 10-minute, and 15-minute) are also investigated. Several important conclusions can be drawn based on the results. First, the periodic functions can improve the prediction performance of machine learning models for more than 60 minutes ahead prediction and improve the over 30 minutes ahead prediction accuracy of all statistical models. Second, the considered three periodic functions have slight difference in improving prediction accuracy of the six prediction models during multistep ahead prediction. Third, with the increase of prediction steps, the impact of periodic function on the prediction model becomes obvious. Fourth, for the same prediction step, the effect of periodic function is more

obvious with the increase of data aggregation level. For future work, since nonrecurrent events (incidents, special events, etc.) may disturb the cyclical pattern of travel time, it will be interesting to analyze and compare the impacts of periodic functions on prediction models under nonrecurrent traffic conditions. In addition, artificial intelligence has greatly promoted the development of traffic science. Especially deep learning algorithms, such as deep residual networks, cyclic neural networks and convolutional neural networks, have been rapidly developed in transportation field. It is also interesting to examine the impact of different periodic functions on deep learning algorithms.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was funded by the National Key Research and Development Program of China (grant no. 2018YFE0102800).

References

- [1] 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.
- [2] J. Xia, M. Chen, and W. Huang, "A multistep corridor travel-time prediction method using presence-type vehicle detector data," *Journal of Intelligent Transportation Systems*, vol. 15, no. 2, pp. 104–113, 2011.
- [3] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: overview of objectives and methods," *Transport Reviews*, vol. 24, no. 5, pp. 533–557, 2004.
- [4] M. Elhenawy and H. A. Rakha, *Expected Travel Time and Reliability Prediction Using Mixture Linear Regression*, Transportation Research Board, Washington DC, USA, 2016.
- [5] X. Zhang and J. A. Rice, "Short-term travel time prediction," *Transportation Research Part C: Emerging Technologies*, vol. 11, no. 3-4, pp. 187–210, 2003.
- [6] A. Guin, "Travel time prediction using a seasonal autoregressive integrated moving average time series model," in *Proceedings of the 2006 IEEE Intelligent Transportation Systems Conference*, September 2006.
- [7] S. Ishak and H. Al-Deek, "Performance evaluation of short-term time-series traffic prediction model," *Journal of Transportation Engineering*, vol. 128, no. 6, pp. 490–498, 2002.
- [8] Z. Tongyu, K. Xueping, L. Weifeng, Z. Yuan, and D. Bowen, "Travel time prediction for float car system based on time series," in *Proceedings of the 2010 12th International Conference on Advanced Communication Technology (ICACT)*, IEEE, Phoenix Park, South Korea, February 2010.
- [9] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.
- [10] Q. F. Yang, X. W. Wei, C. Y. Lin, Z. L. Li, and X. Y. Liu, "Reliability prediction of travel time based on spatio-temporal

- bayesian model,” *Journal of South China University of Technology (Natural Science Edition)*, vol. 44, pp. 115–122, 2016.
- [11] S. I. Chien and C. M. Kuchipudi, “Dynamic travel time prediction with real-time and historic data,” *Journal of Transportation Engineering*, vol. 129, pp. 609–616, 2003.
- [12] L. Chu, J.-S. Oh, and W. Recker, *Adaptive Kalman Filter Based Freeway Travel Time Estimation*, Transportation Research Board, Washington DC, USA, 2005.
- [13] W.-C. Hong, “Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm,” *Neurocomputing*, vol. 74, no. 12–13, pp. 2096–2107, 2011.
- [14] W. Chun-Hsin, H. Jan-Ming, and D. T. Lee, “Travel-time prediction with support vector regression,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, pp. 276–281, 2004.
- [15] K. L. Li, Y. H. Ma, Y. M. Tian, and J. Xie, “An improved LS-SVR ensemble learning in internet traffic prediction,” *Applied Mechanics and Materials*, vol. 121–126, pp. 3794–3798, 2011.
- [16] J. W. C. V. Lint, “Reliable real-time framework for short-term freeway travel time prediction,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 132, pp. 921–932, 2006.
- [17] J. W. C. V. Lint, “Online learning solutions for freeway travel time prediction,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, pp. 38–47, 2008.
- [18] J. Y. Wang, K. I. Wong, and Y. Y. Chen, “Short-term travel time estimation and prediction for long freeway corridor using NN and regression,” in *Proceedings of the 2012 15th International IEEE Conference on Intelligent Transportation Systems*, September 2012.
- [19] X. Zeng and Y. Zhang, “Development of recurrent neural network considering temporal-spatial input dynamics for freeway travel time modeling,” *Computer-aided Civil and Infrastructure Engineering*, vol. 28, no. 5, pp. 359–371, 2013.
- [20] J. Wang, I. Tsapakis, and C. Zhong, “A space-time delay neural network model for travel time prediction,” *Engineering Applications of Artificial Intelligence*, vol. 52, pp. 145–160, 2016.
- [21] X. Yang, Y. Zou, J. Tang, J. Liang, and M. Ijaz, “Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models,” *Journal of Advanced Transportation*, vol. 2020, Article ID 9628957, 16 pages, 2020.
- [22] J. Tang, L. Zheng, C. Han et al., “Statistical and machine-learning methods for clearance time prediction of road incidents: a methodology review,” *Analytic Methods in Accident Research*, vol. 27, Article ID 100123, 2020.
- [23] Y. Zou, X. Zhu, Y. Zhang, and X. Zeng, “A space-time diurnal method for short-term freeway travel time prediction,” *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 33–49, 2014.
- [24] N. Zou, J. Wang, and G. L. Chang, “A reliable hybrid prediction model for real-time travel time prediction with widely spaced detectors,” in *Proceedings of the 2008 11th International IEEE Conference on Intelligent Transportation Systems*, October 2008.
- [25] C. M. Kuchipudi and S. I. J. Chien, “Development of a hybrid model for dynamic travel-time prediction,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1855, no. 1, pp. 22–31, 2003.
- [26] A. Stathopoulos, L. Dimitriou, and T. Tsekeris, “Fuzzy modeling approach for combined forecasting of urban traffic flow,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 7, pp. 521–535, 2008.
- [27] E. I. Vlahogianni, “Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics,” *Journal of Intelligent Transportation Systems*, vol. 13, no. 2, pp. 73–84, 2009.
- [28] H. Jiang, Y. Zou, S. Zhang, J. Tang, and Y. Wang, “Short-term speed prediction using remote microwave sensor data: machine learning versus statistical model,” *Mathematical Problems in Engineering*, vol. 2016, Article ID 9236156, 13 pages, 2016.
- [29] Y. Zou, X. Hua, Y. Zhang, and Y. Wang, “Hybrid short-term freeway speed prediction methods based on periodic analysis,” *Canadian Journal of Civil Engineering*, vol. 42, no. 8, pp. 570–582, 2015.
- [30] J. Tang, H. Wang, Y. Wang, X. Liu, and F. Liu, “Hybrid prediction approach based on weekly similarities of traffic flow for different temporal scales,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2443, no. 1, pp. 21–31, 2014.
- [31] C. Chen, Y. Wang, L. Li, J. Hu, and Z. Zhang, “The retrieval of intra-day trend and its influence on traffic prediction,” *Transportation Research Part C: Emerging Technologies*, vol. 22, pp. 103–118, 2012.
- [32] D. Park, L. R. Rilett, B. J. Gajewski, C. H. Spiegelman, and C. Choi, “Identifying optimal data aggregation interval sizes for link and corridor travel time estimation and forecasting,” *Transportation*, vol. 36, no. 1, pp. 77–95, 2009.
- [33] E. Vlahogianni and M. Karlaftis, “Temporal aggregation in traffic data: implications for statistical characteristics and model choice,” *Transportation Letters*, vol. 3, no. 1, pp. 37–49, 2011.
- [34] F. Qiao, X. Wang, and L. Yu, *Optimizing Aggregation Level for Its Data Based on Wavelet Decomposition*, Transportation Research Board, Washington D.C., 2003.
- [35] X. Liu, X. Fang, Z. Qin, C. Ye, and M. Xie, “A short-term forecasting algorithm for network traffic based on Chaos theory and SVM,” *Journal of Network and Systems Management*, vol. 19, no. 4, pp. 427–447, 2011.
- [36] Y. Kamarianakis, H. Oliver Gao, and P. Prastacos, “Characterizing regimes in daily cycles of urban traffic using smooth-transition regressions,” *Transportation Research Part C: Emerging Technologies*, vol. 18, no. 5, pp. 821–840, 2010.