This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

An Improved Fuzzy Neural Network for Traffic Speed Prediction Considering Periodic Characteristic

Jinjun Tang, Fang Liu, Yajie Zou, Weibin Zhang, and Yinhai Wang

***Abstract*— This paper proposes a new method in construction fuzzy neural network to forecast travel speed for multi-step ahead based on 2-min travel speed data collected from three remote traffic microwave sensors located on a southbound segment of a fourth ring road in Beijing City. The first-order Takagi–Sugeno system is used to complete the fuzzy inference. To train the evolving fuzzy neural network (EFNN), two learning processes are proposed. First, a *K*-means method is employed to partition input samples into different clusters and a Gaussian fuzzy membership function is designed for each cluster to measure the membership degree of samples to the cluster centers. As the number of input samples increases, the cluster centers are modified and membership functions are also updated. Second, a weighted recursive least squares estimator is used to optimize the parameters of the linear functions in the Takagi–Sugeno type fuzzy rules. Furthermore, a trigonometric regression function is introduced to capture the periodic component in the raw speed data. Specifically, the predicted performance between the proposed model and six traditional models are compared, which are artificial neural network, support vector machine, autoregressive integrated moving average model, and vector autoregressive model. The results suggest that the prediction performances of EFNN are better than those of traditional models due to their strong learning ability. As the prediction time step increases, the EFNN model can consider the periodic pattern and demonstrate advantages over other models with smaller predicted errors and slow raising rate of errors.**

***Index Terms*— Speed prediction, evolving fuzzy neural network, K-means clustering, remote traffic microwave sensors.**

I. Introduction

Accurate prediction of traffic information is a key step to achieve the performance of Intelligent Transportation System (ITS), especially in Advance Traffic Management System (ATMS) and Advanced Traveler Information Systems (ATIS) [1]–[5]. Using the forecasted information, such as traffic volume data, travel time data and traffic condition information, travelers can re-plan the traveling paths to save their time and cost. Furthermore, transportation agencies can also improve the efficiency of management in traffic system based on forecasted information. Travel speed is an important indicator to estimate traffic conditions in road networks. Comparing with general collecting approaches, loop detectors and GPS equipments, Remote Traffic Microwave Sensors (RTMS) [1], [6] is another important non-intrusive device to directly detect instantaneous travel speed of vehicles. RTMS is installed on the side of the road, and it can directly detect moving or stationary objects without interrupting traffic flow. It can detect traffic volume, occupancy and speed for multiple lanes simultaneously although sometimes in severe environment. As its high measurement accuracy [1], [6] comparing to single loop detector, travel speed data collected from RTMS is used as data source to construct prediction model in this paper.

Short-term traffic flow forecasting models rely on the regularity existing in historical data to predict the traffic patterns in future time periods. A good prediction algorithm usually requires advanced technologies and computational ability to capture high dimensional and nonlinear characteristics in traffic flow data. In the past few years, a large amount of algorithms have been proposed to address traffic prediction problems. Vlahogianni *et al.* [2] summarized existing short-term traffic predictions algorithms up to 2003. And recently, Vlahogianni *et al.* [3] updated the literature from 2004 to 2013. Van Lint and Van Hinsbergen [4] reviewed existing applications of neural network and artificial intelligence in short-term traffic forecasting and classified prediction models into two major categories: parametric approach and nonparametric approach. Existing traffic prediction algorithms ranges from statistical prediction methods [6]–[11], artificial neural networks [12]–[15], fuzzy-neural networks [16]–[18], support vector regression [19]–[23], Kalman filter theory [24]–[29] and hybrid approaches [18], [30]–[34].

In statistical models (SM), researchers used its advantages of good theoretical interpretability and clear calculation construction to forecast traffic flow. The conventional Vector Autoregressive (VAR) Models [7] considering effect of upstream and downstream can improve prediction performance. Some new algorithms, such as CUSUM (cumulative sum) [5], generalized autoregressive conditional heteroskedasticity (GARCH) [8] and Granger causality [9], all can achieve high prediction accuracy for their strong interpretability. Comparing to SM, due to strong generalization and learn-

Manuscript received September 7, 2015; revised January 6, 2016, July 28, 2016, and December 11, 2016; accepted December 17, 2016. This work was supported by the National Natural Science Foundation of China under Grant 51138003 and Grant 51329801. The Associate Editor for this paper was K. Wang. *(Corresponding author: Yajie Zou.)*

J. Tang is with the School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China (e-mail: jinjuntang@163.com).

F. Liu is with the School of Energy and Traffic Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China (e-mail: rcliufang@163.com).

Y. Zou is with the Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 201804, China (e-mail: zouyajie@gmail.com).

W. Zhang and Y. Wang are with the Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195-2700 USA (e-mail: wbin.zhang@outlook.com; yinhai@uw.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2643005

ing ability as well as adaptability, the artificial neural networks (ANN) has been a widely used method in traffic prediction. In the ANN, the neural network [11] is one of most popular approach to forecast traffic flow time series. Recently, some improve algorithm, such as state-space neural network model [13], Long Short-Term Memory Neural Network (LSTM) [14] have demonstrated their superiority to traditional ones for high computational efficiency and performance. The fuzzy-neural networks (FNN) is another important branch of ANN, which combine fuzzy inference system and network structure of neural network and shows good performance in traffic flow prediction. A fuzzy-neural approach in [16] is used to improve traffic flow prediction accuracy. Furthermore, genetic algorithm is employed in [18] to optimize parameters in an adaptive fuzzy rule-based system for forecasting traffic flow in urban arterial networks. From the other angle to employ prediction, the support vector machine (SVM) maps data into a high-dimensional feature space via a nonlinear relationship and then performs linear regression within this space. The support vector regression (SVR) model has been widely used in travel time [20] and speed [23] prediction. In order to improve the efficiency of parameters optimization in SVM, least squares support vector machines (LS-SVMs) [21] is adopted in some actual applications. For the Kalman filter theory (KFT), it captures regression problem in a state space form by minimizing variance for optimal solution. Some researchers applied this method in traffic flow [25] and travel time perdition [26] and they also discussed the effect of starting network parameters to the prediction performance. Towards improve its online-learning ability, a new extended Kalman filter (EKF) is proposed in [29] to predict highway travel time. As every prediction algorithm has its own advantage and applicable conditions, some hybrid models combing merits of different methods are proposed and applied to improve prediction performance. Genetic algorithm combined with adaptive fuzzy rule-based system [18], neural network model combined with the theory of conditional probability and Bayes' rule [30], and support vector machine combined with statistical and heuristic models [34], etc. All these hybrid models produce higher prediction accuracy than single approaches in traffic flow and travel time forecasting.

Traffic data often demonstrate obvious periodic patterns. Over a 24-hour period in a day, there is generally one or two peak hours with congested traffic condition. By considering periodic features in traffic data, we can not only gain better insights into the data but also improve prediction accuracy. Despite its importance, there are limited studies that consider periodic features in traffic prediction. Dendrinos [35] considered traffic as a combination of periodic components and non-periodic dynamics. Stathopoulos and Karlaftis [36] studied common cyclical components of traffic flow between two successive loop detectors by using spectral analysis. Zhang *et al.* [37] proposed a hybrid approach that uses a trigonometric regression function to model the cyclical patterns of the data and indicate that multi-step ahead prediction results can be improved by considering periodic features of traffic. Tchrakian *et al.* [38] developed a real-time short-term traffic prediction algorithm based on spectral analysis.

According to the reviewed literature, most of traffic prediction models are mainly based on traditional statistical methods and computational intelligence (CI) techniques. These two types of models have their different characteristics. (1) The statistical models can provide good theoretical interpretability with clear calculation construction. (2) CI models use a 'black box' to predict traffic conditions and often lacks a good interpretation of the model. However, comparing with statistical models, CI methods are more flexible with no or little prior assumptions for input variables. In addition, these approaches are more capable of processing outliers, missing and noisy data [39]. (3) Few traditional models classify speed data samples into different clusters or levels and finish learning process according to each cluster or level. As an important approach in CI, fuzzy-neural network (FNN) combines unique features of fuzzy inference and neural network: strong knowledge expression ability and learning ability.

The works and contributions in this study are summarized as following: (1) We improve the FNN model and propose a new learning structure to enhance learning ability. The model contains unsupervised and supervised learning process. In the unsupervised learning process, a K-means method is employed to partition input samples into different clusters, and a Gaussian fuzzy membership function is designed for each cluster to measure the membership degree of samples to the cluster centers. In the supervised learning process, a weighted recursive least squares estimator is used to optimize the parameters of the linear functions in the Takagi-Sugeno type fuzzy rules. (2) We extract periodic features from traffic flow data, and the long term pattern of data can be better captured. A trigonometric regression function is used to fit periodic patterns. The FNN model considering periodic features can improve prediction accuracy in terms of multi-step ahead forecasting. (3) A cross validation approach is applied for selecting optimal key parameters such as the number of clusters, the number of trigonometric polynomials for fitting periodic patterns and the split ratio in learning process. Furthermore, comparisons between the proposed method and some other traditional speed prediction methods such as ARIMA, VAR, ANN and SVM are used to verify the effectiveness of the proposed method in actual scenarios.

II. Data Description

The travel speed data used in the study were collected in 4th ring road in Beijing. The segment selected stretches from Dongfengbei Bridge to Zhaoyang Bridge, and its total length is approximately 2.74 km. This segment experiences significant traffic congestions during peak hours. The speed data were collected from three adjacent stations, which are shown in Fig. 1. The distance between each two adjacent stations is about 1.4 km. Location A represents detector 9053, location B is detector 9054 and location C indicates detector 9055. All three detectors monitor southbound traffic with frequency of 2 minutes in 24 hours a day. The missing data for the three stations are all less than 3%, and historical averaged based data imputation method has been implemented to ensure the selected speed data are appropriate for model validation and

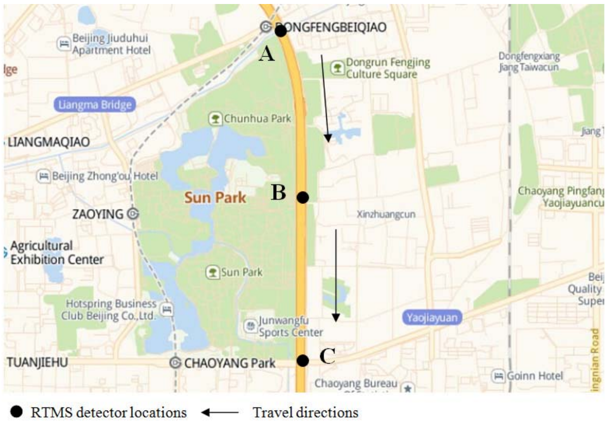


Fig. 1. Three data collection stations in major ring road, Beijing. (The map source comes from Gaode Map).

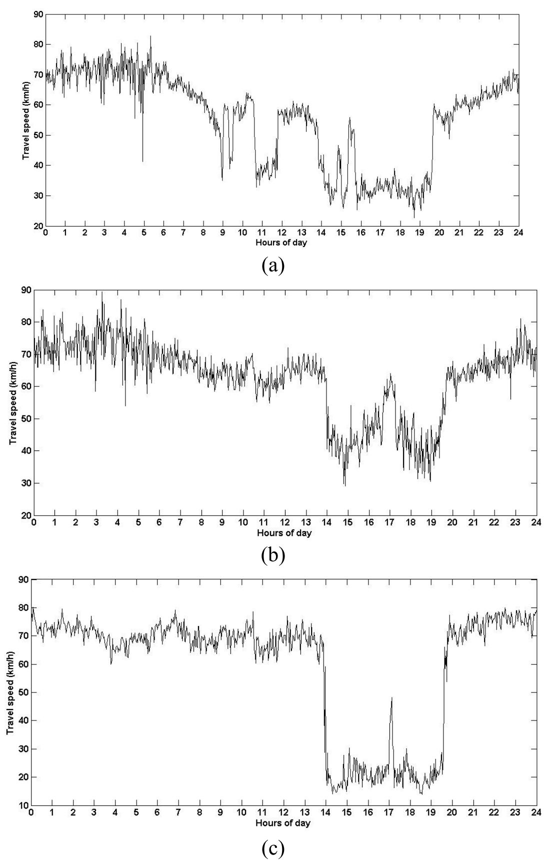


Fig. 2. Travel speed distribution at the three sites in one day. (a) Station A. (b) Station B. (c) Station C.

evaluation in this study. The data collection duration starts from December 1, 2014 to December 31, 2014, total 31 days.

In order to validate the prediction performance of different models and fairly compare the prediction accuracy, data are divided into two parts: training dataset and testing dataset. The data collected from the first 21 days are used to optimize model parameters, and data in last 10 days are employed to validate models effectiveness. The data in one day from each station are plotted in Fig. 2 to show the general trends. For three stations, we can see the speed data demonstrate similar patterns but with different speed values. In addition to the temporal patterns, another important characteristic of speed data is the cyclical pattern. It demonstrates clearly sharp reduction of speed during peak hours and the speed values observed during daytime fluctuate more significantly than the speed values observed at night.

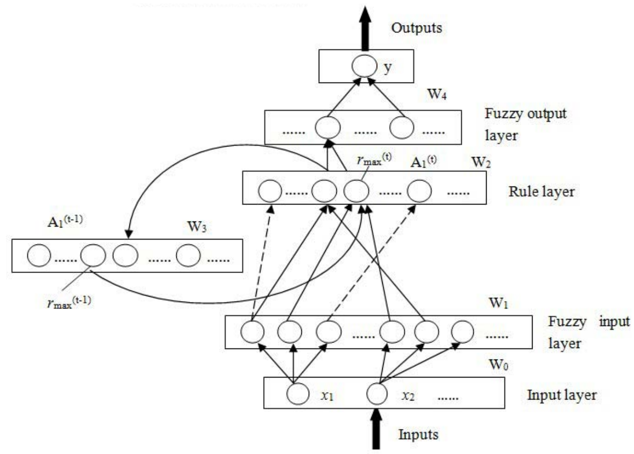


Fig. 3. The architecture of evolving fuzzy neural network with five layers [42].

III. Methodology

A. Evolving Fuzzy Neural System

The evolving fuzzy neural network model (EFNN) was presented in Nikola [40]. An EFNN is an improved structure from the fuzzy neural network (FNN), and it can evolve its structure and functionality from a continuous input data source in an adaptive, life-long, modular way. Furthermore, all nodes in an EFNN are created during the learning process and the nodes representing membership functions can be modified during learning. The structure of an EFNN has five layers as shown in Fig. 3.

The first layer is the input layer, in which the input variables are stored and each node represents a variable. The second layer of nodes quantifies the fuzzy values of the input variables by transforming the input values into membership degrees to which they belong to the membership functions. Each node in the second layer represents a membership function. The number of membership functions and the formulation of each can change during the learning process. In the third layer, the rule nodes can evolve through supervised or unsupervised learning. For this layer, A denotes the activation of the rule nodes, and each rule node r is defined by two vectors of connection weights: $\mathrm{W}_1(r)$ and $\mathrm{W}_2(r)$. The former can be adjusted by unsupervised learning based on similarity measures, and the latter can be adjusted by supervised learning based on the output errors. Between the second and third layers, there is a short-term memory layer which connects to the rule layer and can be used to provide information to it via a feedback loop. The fourth layer of nodes represents fuzzy quantification of the output variables. Finally, the fifth layer represents the real values of the output variables. More details on the structure of an EFNN can be found in [40]–[42].

Comparing to a traditional FNN, an EFNN makes use of improved learning processes, which include two parts: the unsupervised learning process and the supervised learning process. In the unsupervised learning process, the main purpose is to determine parameters in fuzzy variables' membership functions. The supervised learning process is then used to adjust weights in the fuzzy inference system.

1) Clustering Based on K-Means Method: The aim of the K-means method is to classify the l input samples which take the form of m-vectors $x_{\mathrm{i}} = \{x_{\mathrm{i}1}, x_{\mathrm{i}2}, \ldots, x_{il}\}$, $i = 1, 2, \ldots, m$, (here m is the dimension of input data , l is the number of samples) into n clusters, and determine the cluster centers for each cluster under the condition of minimizing an objective function J. The distance between x_{i} and the cluster center c_{j} is first defined in the following equation:

$$d(x_i, c_j) = \sum_{k=1}^{l} \left| x_{ik} - c_j \right| \tag{1}$$

where, $|\cdot|$ represents the general Euclidean distance. Then the objective function is defined as:

$$J = \sum_{i=1}^{m} \sum_{j=1}^{n} d(x_i, c_j) \tag{2}$$

The algorithm for determining the cluster centers with the K-means clustering method can be divided into three processes. First, initialize the cluster center c_{j}. Second, iteratively modify the partition to reduce the sum of the distances for each sample from the centers of the cluster to which the samples belongs. Finally, the process terminates if one of following conditions is satisfied: the value of objective function is below a certain tolerance; the difference in the values of the objective function between adjacent iterations is less than a preset threshold; or the iteration process is complete. As the cluster number can definitely influence the prediction results, we will discuss it in the later sections.

2) Structure of Fuzzy Inference System: In the EFNN, a Takagi-Sugeno type fuzzy inference system is used to construct fuzzy rules. As each sample, $x = [x_1, x_2, \ldots, x_m]$, has n memberships describing the degree to which it belongs to each cluster, the number of rules is equal to the number of clusters K. The rules are shown as follows:

$$\begin{cases} \text{if } x_1 \ is\ R_{11} \text{ and } x_2 \text{ is } R_{12} \text{ and} \ldots \text{and } x_{\mathrm{m}} \text{ is } R_{1\mathrm{m}}, \\ \text{then } y \text{is } f_1(x_1, x_2, \ldots, x_{\mathrm{m}}) \\ \text{if } x_1 \text{ is } R_{21} \text{ and } x_2 \text{ is } R_{22} \text{ and } \ldots \text{ and } x_{\mathrm{m}} \text{ is } R_{2\mathrm{m}}, \\ \text{then } y \\ \text{is } f_2(x_1, x_2, \ldots, x_{\mathrm{m}}) \ldots \ \ldots \\ \text{if } x_1 \text{ is } R_{\mathrm{K}1} \text{ and } x_2 \text{ is } R_{\mathrm{K}2} \text{ and } \ldots \text{ and } x_{\mathrm{m}} \text{ is } R_{\mathrm{Km}}, \\ \text{then} y \text{ is } f_{\mathrm{K}}(x_1, x_2, \ldots, x_{\mathrm{m}}) \end{cases}$$

where R_{ij} indicates a fuzzy set defined by its membership function, x_{j} is the antecedent variable, and f_{i} is the inference consequence of variable y when the ith rule is employed. In this study, the fuzzy membership functions were selected to be of the Gaussian type with two parameters defined as follows:

$$mf(x) = e^{-\frac{(x-\mu)^2}{\sigma^2}} \tag{3}$$

where, *mf* is defined as membership function, μ is the value of the cluster center on the x dimension, σ^2 is the variance of the distance between input samples and the cluster center on the x dimension. Overall, the total number of membership functions is $n \times m$. In the model, a first-order Takagi-Sugeno is used in fuzzy inference system, which means the function f_{i} $(x_1, x_2, \ldots, x_{\mathrm{m}})$, i=1,2,... K, is a linear function. So, for an input data point $x^0 = [x_1^0, x_2^0, \ldots, x_m^0]$, the inferring results of the system, y^0, can be calculated as the weighted average of outputs from each rule:

$$y^0 = \frac{\sum_{i=1}^{K} w_i \cdot f_i(x_1^0, x_2^0, ..., x_m^0)}{\sum_{i=1}^{K} w_i} \tag{4}$$

where, $w_i = \prod_{j=1}^{m} mf_{R_{ij}}(x_j^0)$; $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$.

In the learning process, a least squares estimator (LSE) in [43] and [44] is used to train the linear functions. Each of the linear function can be described as follows:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m \tag{5}$$

The training dataset included p data pairs, $\{([x_{\mathrm{i}1}, x_{\mathrm{i}2}, \ldots, x_{\mathrm{im}}], y_{\mathrm{i}}), i = 1, 2, \ldots, p\}$, and was used to calculate the coefficients $a = [\alpha_0 \ \ \alpha_1 \alpha_2 \ldots \alpha_{\mathrm{m}}]^{\mathrm{T}}$ via the following equation based on LSE:

$$a = (\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}} y \tag{6}$$

where

$$\mathrm{A} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{p1} & x_{p2} & \cdots & x_{pm} \end{pmatrix}$$

and

$$y = [y_1 \quad y_2 \quad y_3 \ldots y_{\mathrm{p}}]^{\mathrm{T}}$$

Furthermore, an improving weighted least squares estimation method in [43] and [44] is used to optimize the parameters.

$$a_w = (\mathrm{A}^{\mathrm{T}}\mathrm{W}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}\mathrm{W} y \tag{7}$$

where

$$\mathrm{W} = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & w_p \end{pmatrix}$$

w_{j} represents the distance between the jth sample and the corresponding cluster center, $j = 1, 2, \ldots, p$. Equation (7) can be rewritten according to the following:

$$\begin{cases} \mathrm{P}_w = (\mathrm{A}^{\mathrm{T}}\mathrm{W}\mathrm{A})^{-1} \\ a_w = \mathrm{P}_w \mathrm{A}^{\mathrm{T}}\mathrm{W} y \end{cases} \tag{8}$$

Define the kth row vector of matrix A in equation (6) to be $b_{\mathrm{k}}^{\mathrm{T}} = [1 x_{\mathrm{k}1} \ x_{\mathrm{k}2} \ldots x_{\mathrm{km}}]$ and denote the kth element of y as y_{k}. Then the vector of coefficient a can be iteratively calculated by equation (9) shown in the following. The calculation process uses a recursive, improved weighted LSE method [42], [43]

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

to complete the optimization.

$$\begin{cases} a_{k+1} = a_k + w_{k+1}\mathrm{P}_{k+1}b_{k+1}(y_{k+1} - b_{k+1}^T a_k) \\ \mathrm{P}_{k+1} = \frac{1}{\lambda}(\mathrm{P}_k - \dfrac{w_{k+1}\mathrm{P}_k b_{k+1} b_{k+1}^T \mathrm{P}_k}{\lambda + b_{k+1}^T \mathrm{P}_k b_{k+1}}) \ k = t, t+1, \ldots, p-1 \end{cases} \tag{9}$$

where λ is the forgetting factor and with a value is generally between 0.8-1.0, a_{t} and P_{t} are the initial values of a and P, which can be calculated in equation (8) by using the first t data pairs from the training dataset. Here, the r is defined as the split ratio, if p represents the total number of training samples, then $r * p$ is the number of samples used in first step and $(1-r) * p$ indicates the number of samples used in second step. Similarly, we will discuss the selection of$r$ in the later sections.

B. Cyclical Patterns

The periodicity of traffic flow in this study means daily similarity. We can observe similar distribution between days (peaks and trough hours) under normal traffic state. So, the length of period T in prediction is one day.

Usually, a periodic function $\gamma(t)$ with period T can be expanded into a Fourier series as follows:

$$\gamma(t) = \sum\nolimits_{k=-\infty}^{+\infty} \beta_k \cdot e^{jk(\frac{2\pi}{T})t} \tag{10}$$

where, β_{k} is the coefficient, and it can be defined as:

$$\beta_k = \frac{1}{T}\int_T \varphi(t) \cdot e^{-jk(\frac{2\pi}{T})t} dt \tag{11}$$

$$e^{j\theta} = \cos\theta + j \cdot \sin\theta \tag{12}$$

Therefore, the Fourier series can be described as a trigonometric polynomials series:

$$g(x) = \sum\nolimits_{-\infty}^{+\infty} \beta_{1k} \cdot \cos(kx) + \sum\nolimits_{-\infty}^{+\infty} \beta_{2k} \cdot \sin(kx) \tag{13}$$

To model the periodic component observed in Fig.2, a combination of sinusoids and cosinusoids is used, which is referred to as the trigonometric regression function. This approach can describe regular cyclical patterns or periodic variations and has been used for in various time series data analysis [45]. Using the observed 2-minute average travel speed values, the daily average 2-minute speed at each station is calculated by using $V_t = \frac{1}{21}\sum_{d=1}^{21} v_t^d$, where V_{t} is daily average 2-minute average speed at time t; v_t^d is 2-minute average speed at time t on day d; $t = 1, 2, \ldots, 720$ (As we use daily similarity to express periodicity of traffic flow, the total number of data samples collected in one day is 720); and $d = 1, 2, \ldots 21$ is the number of training days in a month used to estimate the coefficients in equation (14). Considering equation (13), a limited number of trigonometric polynomials with sufficient accuracy is used to represent the period component of the travel speed time series:

$$\begin{aligned} M_u &= m_0 + m_1 \sin(\frac{2\pi u}{720}) + m_2 \cos(\frac{2\pi u}{720}) \\ &\quad + \ldots + m_{2n-1} \sin(\frac{2n\pi u}{720}) + m_{2n} \cos(\frac{2n\pi u}{720}) \end{aligned} \tag{14}$$

where M_{u} is the estimated periodic component at time u, $u = 1, 2, \ldots 720$, n is the number of trigonometric polynomials, $m_0, m_1, \ldots, m_{2\mathrm{n}}$ are the coefficients. For the daily average 2-minute speed, a least squares estimation method is used to determine the parameters in Equation (14). As the number of trigonometric polynomials can affect the prediction performance of the model, the selection of number of trigonometric polynomials will be discussed in the results section.

C. Structure of the Proposed EFNN+CP Prediction Method

As we discussed in Section two, speed values often have a daily periodic pattern. Thus, in this study, original data are divided into two parts.

The hybrid prediction process in the study includes following several steps:

Step1 The training dataset, speed data collected in first 21 days at 2-min time scale, was used in trigonometric regression function to fit daily periodic pattern based on Equation (14);

Step2 Considering raw speed data contain periodic component and residual part:

$$S_t = M_t + S_t^r \tag{15}$$

where S_{t} is the speed at time t at a selected station, M_{t} is the periodic component, S_t^r is the residual part after removing the periodic component. We use raw speed data minus periodic component, and then obtain residual errors. One is the cyclic component and represents the periodic trend of speed, and the other is the irregular component or residual component.

Step3: The residual errors are used as training dataset to optimize parameters in EFNN model, and then predict residual errors in future steps.

Step4: Through combing predicted residual errors and periodic component, the predicted values for real speed data are calculated

IV. Prediction Results and Discussion

A. Evaluation Indicators

For all the prediction algorithms, the speed data in first 21 days are used for training models and data in last 10 days are used for validating models. To evaluate the multi-step prediction performance of different models, three performance measures, the mean absolute error (MAE), the mean absolute percentage error (MAPE) and the root mean square error (RMSE) are considered. The unit of the MAE and RMSE is km per hour. The equations for calculating MAE, MAPE and RMSE are shown as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} \left|\hat{S}_i - S_i\right|}{N} \tag{16}$$

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N} \frac{\left|\hat{S}_i - S_i\right|}{S_i} \times 100\% \tag{17}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\hat{S}_i - S_i)^2}{N}} \tag{18}$$

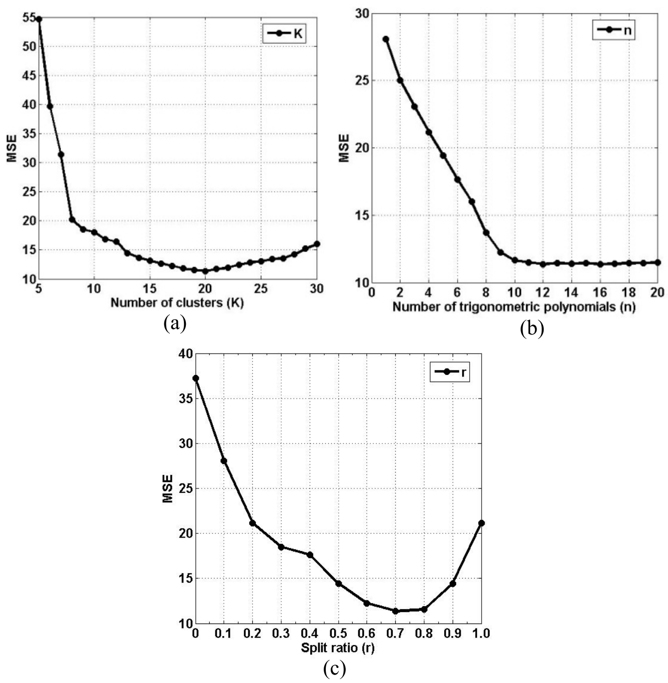


Fig. 4. Parameter selection of model in station A for the one-step-ahead prediction. (a) Effect of K. (b) Effect of n. (c) Effect of r.

where, N is the number of observations, S_i is the actual speed at time i at station, and $\hat{S}_i$ is the predicted speed. Furthermore, in order to further evaluate the performance of the all models, both one-step and multi-step ahead prediction (i.e., 3-step (6 minutes), 5-step (10 minutes) and 10-step (20 minutes)) are considered.

B. Parameter Selection

From the aforementioned introduction, the model in this study contains several key parameters to be determined for prediction: the number of clusters, K; the number of trigonometric polynomials for fitting periodic patterns, n; and the split ratio in learning process, r. (Here, for the forgetting factor λ in Equation (9), its value is generally between 0.8-1.0 as suggested in previous two studies [42], [43]. In this study, we tried different values of λ and found there is no obvious difference in prediction results. So, we used 0.95 when the accuracy reaches its maximum in our study). For K, too small or too large value may cause inaccurate predicted results, thus we assume it ranges from 5 to 30. n ranges from 1 to 20, and r will be selected in the ranges of [0, 1] with an increment by 0.1. In the process of parameter selection, mean square error is used (MSE) as evaluation index. These parameters are determined based on cross validation only using the training dataset to ensure fairness in comparison. Here, Fig. 4 only demonstrates the process of parameter selection in station A for one-step-ahead prediction.

First, the impact of K on the prediction performance is examined while keeping the other two parameters to be constant. Fig. 4a shows the prediction accuracy against the number of cluster K increasing from 1 to 30. As we can see, when the value of K is less than 10 ($K < 10$), the prediction error is high. The reason is that K-means method is unable to effectively divide the input samples into different patterns with small cluster number, which also leads to poor learning effect. When the value of K is large ($K > 20$), the prediction error increases slightly. The reason is that K-means method classifies the input data into sparse patterns with large K, which can significantly reduce the association in raw data. Furthermore, large number of clusters can increase the complexity of the training process by adding the number of parameters, such as the number of fuzzy rules and fuzzy membership function, see equation (4). Finally, a large amount of parameters in Takagi-Sugeno inference would cause over fitting in training process to a certain extent. Based on the above analysis, the number of clusters is selected as 20 in the prediction model.

TABLE I
PARAMETERS OF THREE STATIONS FOR DIFFERENT PREDICTING STEPS

		1 step	3 step	5 step	10 step
Station A	K	20	22	25	10
	n	12	10	15	15
	r	0.7	0.6	0.8	0.6
Station B	K	18	25	20	10
	n	14	12	15	10
	r	0.5	0.8	0.8	0.4
Station C	K	15	13	17	10
	n	10	14	15	16
	r	0.4	0.7	0.6	0.4

Second, for the number of trigonometric polynomials used to fit periodic patterns n, based on the results in Fig. 4b, small n ($n < 10$) will cause high prediction errors. The reason is that trigonometric function with small number of components in equation (14) cannot adequately describe the cyclical patterns observed in speed data. As n increases, ($n > 10$), the prediction performance can be improved significantly. Thus, the $n = 12$ is chosen with lowest MSE.

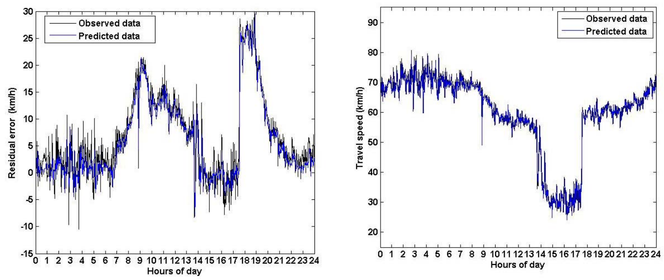
Finally, for the split ratio r, Fig. 4c displays the relationship between r and MSE. When the value of r is too small ($r < 0.4$) or too large ($r > 0.9$), the MSE is high. When the value of r ranges from 0.5 to 0.8, the RMSE is relative low and stable. The reason is that the small value of r can result in inappropriate initial parameters from equation (8), then these initial values are used to update parameters in equation (9), finally the errors will be cumulated and cause inaccurate speed prediction. Similarly, the large value of r will result in the final parameters excessively depend on initial values, and error adjustment mechanism plays only a small role in learning process. In summary, r is set to 0.7, which can not only guarantee proper initial parameters but also make full use of the function of error adjustment.

Similarly, the same approach employed for parameters selection in stations B and C. Table 1 provides all parameters optimized in this study for three stations with different step-ahead prediction.

C. Comparison of Prediction Results

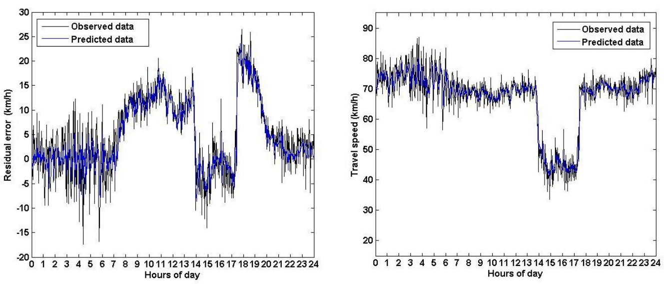
In ANN, Back Propagation Neural Network (BPNN) and Nonlinear Autoregressive with Exogenous Inputs Neural Network (NARXNN) are selected as the candidate models for comparison purposes, and they both have one hidden layer

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) (b)

Fig. 5. Prediction results of speed at station A with one-step ahead using the EFNN+CP model. (Time Scale: 2min). (a) Residual component. (b) Actual values.



(a) (b)

Fig. 6. Prediction results of speed at station B with one-step ahead using the EFNN+CP model. (Time Scale: 2min). (a) Residual component. (b) Actual values.

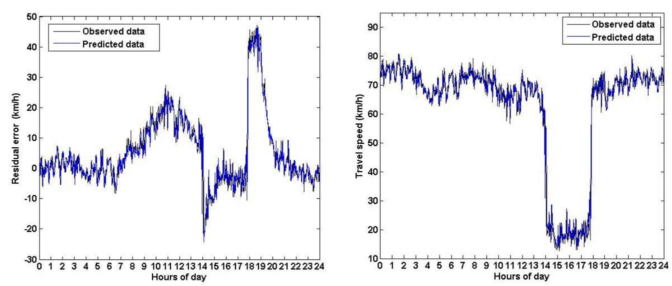
with 50 neurons. Neural network tool in MATLAB is used to optimize parameters. In SVM, Radial Basis Function (RBF) and linear are adopted as kernel functions. For the parameters optimization, [33] provides detailed introduction. The parameters of ARIMA and VAR models are estimated using the maximum likelihood estimation available in forecast and vars packages in R. When forecasting future speed values, the best order of the ARIMA model is determined by the Akaike Information Criterion (AIC) values using the most recent 21 days of speed data. The VAR model is implemented using a maximal order of 10. And the best order of the VAR model is also selected based on the AIC values using the differenced speed data [9], [46]. In order to make a fare comparison, recent training dataset in first 21 days are used to optimize parameters in models and testing dataset in last 10 days are utilized to validate model performance. Furthermore, the dimension of input samples for all models is 10, which means 10 previous speed data are used to predict current values. Tables 2, 3 and 4 provide the MAE, MAPE and RMSE values of different models for different forecasting horizons. Note that in all tables, bold values indicate the smallest MAE, MAPE and RMSE values. Furthermore, the EFNN+CP mean the EFNN model considering cyclical patterns. Fig. 5, 6 and 7 only show the prediction results of proposed models in three stations for one forecasting step comparing with observed speed data in one weekday.

Furthermore, a naive prediction method is applied in model comparison: historical model (HM). In HM, we first use current time period's observations to predict observations in the next time, which is denoted as HM_1. Furthermore, in order to make fair comparison with other models, HM also calculates

TABLE II

PREDICTION ACCURACY OF MODELS FOR DIFFERENT FORECASTING STEPS AHEAD IN STATION A (TIME SCALE: 2min)

MAE	Number of forecasting steps ahead			
	1	3	5	10
EFNN+CP	**2.4782**	**2.7995**	**2.9099**	**3.7218**
EFNN	2.5503	2.9205	3.2768	4.0628
BPNN	2.5027	3.0243	3.4576	4.3553
NARXNN	2.5041	3.0232	3.4482	4.3357
SVM-RBF	2.6537	3.1135	3.4858	4.2625
SVM-LIN	2.7492	3.1160	3.4714	4.2604
ARIMA	2.6777	3.1501	3.5291	4.3233
VAR	2.6835	3.2931	3.7145	4.6030
HM_1	2.9912	4.0150	4.8014	5.6759
HM_10	2.9316	3.5041	4.1644	5.0124
MAPE (%)	Number of forecasting steps ahead			
	1	3	5	10
EFNN+CP	**5.2528**	**5.9654**	**6.8399**	**8.8431**
EFNN	5.3734	6.1237	7.1744	9.0132
BPNN	5.2831	6.4927	7.5472	9.5735
NARXNN	5.2923	6.4995	7.4855	9.6267
SVM-RBF	5.5839	6.6952	7.5901	9.4775
SVM-LIN	5.2878	6.5408	7.6922	9.4937
ARIMA	5.6290	6.7794	7.6974	9.6376
VAR	5.6271	6.7202	8.2556	9.5937
HM_1	6.2070	7.3161	9.1692	10.6664
HM_10	6.0827	7.3986	8.6314	10.1775
RMSE	Number of forecasting steps ahead			
	1	3	5	10
EFNN+CP	**3.3704**	**4.1398**	**4.3292**	**5.3975**
EFNN	3.5224	4.2844	4.7556	5.8504
BPNN	3.4842	4.4918	5.1848	6.4242
NARXNN	3.4729	4.5084	5.2121	6.4153
SVM-RBF	3.6882	4.6544	5.3302	6.5309
SVM-LIN	3.5648	4.5394	5.4574	6.6925
ARIMA	3.7169	4.6993	5.3876	6.6059
VAR	3.7343	4.5941	5.7684	6.7732
HM_1	4.0536	5.2348	6.8475	8.0930
HM_10	4.3017	5.0613	6.1539	7.3311



(a) (b)

Fig. 7. Prediction results of speed at station C with one-step ahead using the EFNN+CP model. (Time Scale: 2min). (a) Residual component. (b) Actual values.

the average values of ten recent observations as predictions in the next or future time period, which is represented as HM_10. The prediction results at deferent forecasting steps in three stations are added in Tables 2, 3 and 4. From the results comparison, the performance of model proposed in this study is obviously improved in three stations, especially for the results predicted in multi-steps ahead. Moreover, the HM_10 is superior to HM_1 for considering more historical information in forecasting process.

Based on the reported values in tables and corresponding figures, several interesting findings can be obtained:

(1) As expected, the prediction accuracy of speed deteriorates as the prediction time steps increase for all models. The results in Tables 2,3 and 4 show that the MAE, MAPE and

TABLE III

PREDICTION ACCURACY OF MODELS FOR DIFFERENT FORECASTING STEPS AHEAD IN STATION B (TIME SCALE: 2min)

MAE	Number of forecasting steps ahead			
	1	3	5	10
EFNN+CP	2.9415	**3.0197**	**3.1449**	**3.2554**
EFNN	3.0041	3.1256	3.2989	3.7554
BPNN	2.9402	3.1513	3.3712	3.8135
NARXNN	**2.9297**	3.1434	3.3591	3.8208
SVM-RBF	2.9825	3.2323	3.4342	3.8567
SVM-LIN	3.1945	3.2111	3.5712	3.8806
ARIMA	3.0144	3.2669	3.4683	3.8959
VAR	3.0244	3.3033	3.7154	4.0356
HM_1	3.6318	4.1459	4.6554	5.1982
HM_10	3.4668	3.8968	4.0939	4.5207
MAPE (%)	Number of forecasting steps ahead			
	1	3	5	10
EFNN+CP	4.9603	**5.0842**	**5.2497**	**5.8356**
EFNN	5.0135	5.2328	5.6855	6.4528
BPNN	4.9518	5.3793	5.7737	6.6007
NARXNN	**4.9325**	5.3470	5.7459	6.6096
SVM-RBF	4.9811	5.4445	5.8253	6.6268
SVM-LIN	5.4039	5.4243	6.2740	6.8330
ARIMA	5.0383	5.5078	5.8899	6.7010
VAR	5.0470	5.6490	6.4091	6.8736
HM_1	6.0168	6.5405	7.1937	7.5408
HM_10	5.7682	6.4879	6.8654	7.6689
RMSE	Number of forecasting steps ahead			
	1	3	5	10
EFNN+CP	4.0459	**4.1264**	**4.3497**	**4.8412**
EFNN	4.0651	4.2819	4.5633	5.1984
BPNN	4.0190	4.3947	4.7273	5.4130
NARXNN	**3.9993**	4.3687	4.7069	5.4215
SVM-RBF	4.0932	4.4750	4.8115	5.4978
SVM-LIN	4.3078	4.4691	4.9729	5.5611
ARIMA	4.1307	4.5213	4.8630	5.5669
VAR	4.1455	4.6963	5.1365	5.7966
HM_1	4.9670	5.4554	5.8980	6.6445
HM_10	4.5136	5.3595	5.6699	6.3290

TABLE IV

PREDICTION ACCURACY OF MODELS FOR DIFFERENT FORECASTING STEPS AHEAD IN STATION C (TIME SCALE: 2min)

MAE	Number of forecasting steps ahead			
	1	3	5	10
EFNN+CP	2.5301	**3.0442**	**3.6307**	**4.7890**
EFNN	**2.5244**	3.2733	4.0606	5.4660
BPNN	2.5694	3.4613	4.1736	5.6407
NARXNN	2.5484	3.4757	4.1458	5.7245
SVM-RBF	2.6617	3.6111	4.2583	5.6255
SVM-LIN	2.8896	3.7087	4.3468	5.6221
ARIMA	2.6675	3.6235	4.2857	5.6745
VAR	2.6611	3.6798	4.3051	5.4380
HM_1	3.3218	4.4472	4.9699	6.2276
HM_10	2.9436	3.9955	4.5117	5.9373
MAPE (%)	Number of forecasting steps ahead			
	1	3	5	10
EFNN+CP	6.6320	**7.5933**	**10.2586**	**12.7613**
EFNN	6.5312	8.3345	11.1587	14.4796
BPNN	6.5606	8.9994	11.3564	16.3149
NARXNN	**6.4875**	9.1176	11.2402	16.6231
SVM-RBF	6.7398	9.3483	11.5343	16.4376
SVM-LIN	7.0707	10.1794	10.9789	15.5072
ARIMA	6.7822	9.4230	11.6734	16.7901
VAR	6.6898	9.1819	11.0211	14.5708
HM_1	8.3646	10.8014	12.7981	17.4989
HM_10	7.9010	10.2233	12.0328	16.7715
RMSE	Number of forecasting steps ahead			
	1	3	5	10
EFNN+CP	3.8267	**5.2468**	**6.5848**	**8.7186**
EFNN	**3.8121**	5.6946	7.0398	9.2574
BPNN	3.8802	5.9716	7.3791	9.8210
NARXNN	3.8909	5.9325	7.3591	9.8298
SVM-RBF	3.9976	6.1269	7.5199	9.9518
SVM-LIN	4.5331	6.1465	8.2970	10.4241
ARIMA	4.0164	6.1734	7.5708	10.0179
VAR	4.0014	6.2533	7.7190	10.3192
HM_1	5.8897	7.4753	9.1064	11.5512
HM_10	4.6649	6.7879	8.6229	10.6186

RMSE values for 10-step ahead forecasting are significantly larger than the results of one-step ahead, and the prediction accuracy and stability of multi-step ahead decrease comparing to the results of one-step-ahead prediction.

(2) The EFNN method obtained better prediction performance than the ANN, SVM and statistical models due to its learning ability. In the EFNN, the input samples are classified into different clusters to reflect variation of traffic flow patterns. In the training process, the speed data in different patterns are considered as input variables and future data as the output variable. With the aim to best describe different patterns, parameters are optimized to accurately describe the non-linear relationship between input and output variables. Thus, the unsupervised and supervised learning abilities of the EFNN are the main advantages comparing to other six traditional models.

(3) For the multi-step ahead forecasting, the model considering periodic characteristics gradually show its advantage. As the time step increases, the difference in prediction performance among models becomes larger, whereas the EFNN with CP model can consistently provide the lowest MAE, MAPE and RMSE values. For example, if we observe the RMSE values for five-step-ahead prediction in station A, see Table 2, the proposed model can improve the prediction results by 20% and 23% when comparing with BPNN and SVM-RBF models. For the other two stations, the EFNN with CP model also produces lowest errors, which demonstrates that prediction accuracy, especially for the multi-step-ahead, can be improved obviously by considering cyclical patterns in raw data.

(4) When comparing the results between machine learning and statistical model, the BPNN, NARXNN and SVM-RBF can clearly outperform two traditional statistical models: ARIMA and VAR. The reason is that these three machine learning models have complex structure and strong learning ability. SVM-LIN produces similar prediction results with ARIMA and VAR. For the models in machine learning, the prediction performance of ANN is superior to that of SVM. Furthermore, the BP and NARXNN both demonstrate similar accurate prediction performance. In the SVM, as RBF is more flexible than linear function, the SVM-RBF can clearly outperform SVM-LIN.

Furthermore, in order to deeply discuss the function of periodic characteristic, we further compare predicted accuracy between BPNN (it has relative high accuracy comparing with other traditional models) and proposed model for different predicting step in three stations, from one step (2 min ahead) to thirty steps (60 min ahead). Fig.8 shows the difference of RMSE enhances gradually as the prediction steps increase from 1 to 30. Although the predicted results are close for small prediction steps (step <10), for large prediction steps, the fitted trigonometric polynomials play a key role in describing the

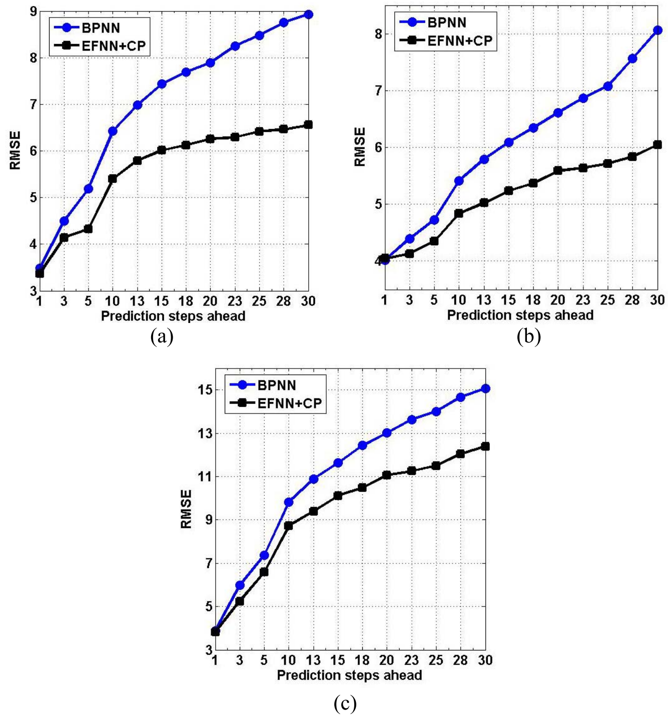


Fig. 8. Predicted results for multi-step ahead. (Time Scale: 2min). (a) Station A. (b) Station B. (c) Station C.

periodic pattern in speed data and thus improve the prediction performance of the models proposed in this study.

In order to explore more details in periodic pattern, we divide data set into two parts: workdays and weekends. Furthermore, for the dataset collected in workdays and weekends, we use trigonometric function to fit the periodic patterns respectively. Table 5 shows the prediction results of different methods (EFNN+CP, BPNN, SVM-RBF, ARIMA) in workdays and weekends. Here, we only calculate prediction errors RMSE for 1-step and 3-step ahead. Comparing with the results in Table 1, Table 2 and Table 3, the prediction accuracy increases by average 5%.

V. Conclusions

In this paper, a new travel speed prediction method is developed based on an evolving fuzzy neural network by using traffic speed data collected from three RTMS located on 4th ring road Beijing City. This EFNN model is based on the Takagi-Sugeno fuzzy inference system, in which n fuzzy rules are activated to calculate the output vectors for a given set of input vectors. In the training process, a K-means method was first used to cluster the input vectors and calculate the cluster centers. Then, the Gaussian-type fuzzy membership functions were designed to evaluate the membership degree of the input samples to each of the clusters. Finally, an improved weighted recursive LSE was used to optimize the parameters in the linear Takagi-Sugeno fuzzy inference functions. In the models performance comparison, six traditional methods are selected as candidates: BPNN, NARXNN, SVM-RBF, SVM-LIN, ARIMA and VAR. The performance of the proposed method was evaluated in terms of three measurement criteria: MAE, RMSE, and MAPE. Based on the results from this

TABLE V

Prediction Accuracy of Models in Three Stations According to RMSE (Time Scale: 2min)

Station	Days of Week	Predict Steps	Imputation Methods			
			EFNN +CP	BPNN	SVM -RBF	ARIMA
A	workday	1 step	3.2947	3.3906	3.5600	3.6117
		3 step	3.8700	4.2173	4.3859	4.4303
	weekend	1 step	3.3188	3.4299	3.6379	3.6605
		3 step	3.9128	4.2721	4.4416	4.4743
B	workday	1 step	3.8522	3.8474	3.9074	3.9302
		3 step	3.9262	4.1931	4.2670	4.3159
	weekend	1 step	3.9031	3.8778	3.9476	3.9828
		3 step	3.9788	4.2310	4.3065	4.3500
C	workday	1 step	3.6505	3.7197	3.8077	3.8350
		3 step	4.8670	5.5787	5.7174	5.7528
	weekend	1 step	3.6970	3.7473	3.8577	3.8754
		3 step	4.9199	5.6330	5.7592	5.8029

study, several interesting conclusions could be drawn. First, the prediction accuracy of speed decreased as the prediction time steps increased for all models. Second, the EFNN was superior to ANN, SVM and SM models. Third, the EFNN considering CP model could significantly improve the multi-step ahead prediction accuracy. Fourth, machine learning models outperformed two traditional statistical models. Among the machine learning models, the prediction performance of ANN was better than that of SVM.

For future studies, we will improve current research from following several aspects:

(1) Consider the impact of traffic event on the speed prediction performance. Non-recurrent events (e.g., incidents, special events, etc.) can disturb the periodic pattern of speed. Thus, it is useful to develop a regime-switching EFNN prediction model to take into account the alternating no-congested and congested traffic regimes.

(2) Include the spatial-temporal features of speed in the prediction model. A lot of works have demonstrated that speed variation patterns at one station show high spatial and temporal correlation with that at adjacent stations. This spatial-temporal property can be used to modify the prediction patterns of speed in a specific location from other related locations.

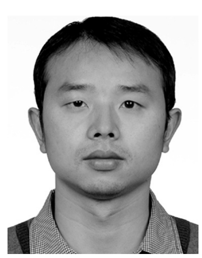
(3) In traffic flow analysis, three variables are generally used: traffic volume, occupancy and speed to describe its state transition. Furthermore, in different states, these three variables express different correlation in Macroscopic Fundamental Diagram. In the further research, volume or occupancy can be used as inputs variables of the proposed model to improve prediction accuracy.

(4) The speed prediction for arterial roads or networks becomes more applicable in traffic management and control. It is useful to quickly respond to some special events and improve managing efficiency by perceiving future patterns of speed in arterial or local network based on accurate prediction results. One practical future research will focus on how to apply the proposed method from a single station to roads or network level.

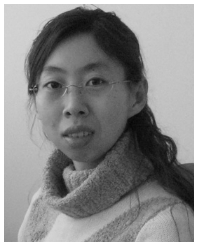
References

[1] X. Yu and P. D. Prevedouros, “Performance and challenges in utilizing non-intrusive sensors for traffic data collection,” *Adv. Remote Sens.*, vol. 2, no. 2, pp. 45–50, 2013.

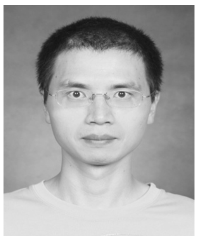
[2] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: Overview of objectives and methods," *Transp. Rev.*, vol. 24, no. 5, pp. 533–557, 2004.
[3] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.
[4] J. W. C. Van Lint and C. P. I. J. Van Hinsbergen, "Short term traffic and travel time prediction models," in *Transportation Research Circular*, R. Chowdhury and S. Sadek, Eds. Washington, DC, USA: National Academy Press, Nov. 2012.
[5] M. Cetin and G. Comert, "Short-term traffic flow prediction with regime switching models," *Transp. Res. Rec.*, vol. 1965, pp. 23–31, Apr. 2006.
[6] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.
[7] P. Shang, X. Li, and S. Kamae, "Chaotic analysis of traffic time series," *Chaos, Solitons Fractals*, vol. 25, pp. 121–128, Jul. 2005.
[8] S. R. Chandra and H. Al-Deek, "Cross-correlation analysis and multivariate prediction of spatial time series of freeway traffic speeds," *Transp. Res. C, Emerg. Technol.*, vol. 2061, pp. 64–76, Sep. 2008.
[9] Y. Zhang, A. Haghani, and X. Zeng, "Component GARCH models to account for seasonal patterns and uncertainties in travel-time prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 719–729, Feb. 2015.
[10] L. Li, X. Su, Y. Wang, Y. Lin, Z. Li, and Y. Li, "Robust causal dependence mining in big data network and its application to traffic flow predictions," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 292–307, Sep. 2015.
[11] Y. Zou, X. Hua, Y. Zhang, and Y. Wang, "Hybrid short-term freeway speed prediction methods based on periodic analysis," *Can. J. Civil Eng.*, vol. 42, no. 8, pp. 570–582, 2015.
[12] Q. Ye, W. Y. Szeto, and S. C. Wong, "Short-term traffic speed forecasting based on data recorded at irregular intervals," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1727–1737, Dec. 2012.
[13] X. Ma, H. Yu, Y. Wang, and Y. Wang, "Large-scale transportation network congestion evolution prediction using deep learning theory," *PLoS ONE*, vol. 10, no. 3, p. e0119044, 2015.
[14] J. W. C. Van Lint, S. P. Hoogendoorn, and H. J. Van Zuylen, "Accurate freeway travel time prediction with state-space neural networks under missing data," *Transp. Res. C, Emerg. Technol.*, vol. 13, pp. 347–369, Oct. 2005.
[15] S. Ghosh-Dastidar and H. Adeli, "Wavelet-clustering-neural network model for freeway incident detection," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 18, no. 5, pp. 325–338, 2003.
[16] H. Yin, S. C. Wong, J. Xu, and C. K. Wong, "Urban traffic flow prediction using a fuzzy-neural approach," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 2, pp. 85–98, 2002.
[17] Q. Chai, M. Pasquier, and B. B. S. Lim, "POP-TRAFFIC: A novel fuzzy neural approach to road traffic analysis and prediction," in *Proc. IEEE Intell. Transp. Syst.*, Sep. 2006, pp. 133–146.
[18] L. Dimitriou, T. Tsekeris, and A. Stathopoulos, "Adaptive hybrid fuzzy rule-Based system approach for modeling and predicting urban traffic flow," *Transp. Res. C, Emerg. Technol.*, vol. 16, pp. 554–573, Oct. 2008.
[19] S. H. Lam and T. D. Toan, "Short-term travel time predictions using support vector regression," in *Proc. Annu. Meeting Transp. Res. Board*, 2008, paper 08-0670.
[20] C. H. Wu, J. M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Trans. Intell. Transp. Syst.*, vol. 125, no. 6, pp. 515–523, Jun. 2004.
[21] Y. Zhang and Y. Liu, "Traffic forecasting using least squares support vector machines," *Transportmetrica*, vol. 5, no. 3, pp. 193–213, 2009.
[22] Y. Zhang and Y. Xie, "Forecasting of short-term freeway volume with υ-support vector machines," *Transp. Res. Rec.*, vol. 2024, pp. 92–99, Aug. 2007.
[23] M. T. Asif *et al.*, "Spatiotemporal patterns in large-scale traffic speed prediction," *IEEE Transp. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 794–804, Feb. 2014.
[24] Y. Zhao, Y. Liu, L. Shan, and B. Zhou, "Dynamic analysis of Kalman filter for traffic flow forecasting in sensor nets," *Inf. Technol. J.*, vol. 11, pp. 1508–1512, Oct. 2012.
[25] H. Chen and S. Grant-Muller, "Use of sequential learning for short-term traffic flow forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 9, pp. 319–336, Sep. 2001.
[26] S. I. J. Chien and C. M. Kuchipudi, "Dynamic travel time prediction with real-time and historic data," *J. Transp. Eng.*, vol. 129, no. 6, pp. 608–616, 2003.
[27] S. Chien, X. Liu, and K. Ozbay, "Predicting travel times for the South Jersey real-time motorist information system," *Transp. Res. Rec.*, vol. 1855, pp. 32–40, Jan. 2003.
[28] J. W. C. Van Lint, "Online learning solutions for freeway travel time prediction," *IEEE Transp. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 38–47, Jan. 2008.
[29] Y. Wang, M. Papageorgiou, and A. Messmer, "RENAISSANCE—A unified macroscopic model-based approach to real-time freeway network traffic surveillance," *Transp. Res. C, Emerg. Technol.*, vol. 14, no. 3, pp. 190–212, 2006.
[30] W. Zheng, D.-H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: Bayesian combined neural network approach," *J. Transp. Eng.*, vol. 132, pp. 114–121, Sep. 2006.
[31] K. Hamad, M. T. Shourijeh, E. Lee, and A. Faghri, "Near-term travel speed prediction utilizing Hilbert–Huang transform," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 24, pp. 551–576, Nov. 2009.
[32] Y. Xie, Y. Zhang, and Z. Ye, "Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 22, pp. 326–334, Sep. 2007.
[33] 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," *Transp. Res. Rec.*, vol. 2443, pp. 21–31, Oct. 2014.
[34] C. Dong, S. H. Richards, Q. Yang, and C. Shao, "Combining the statistical model and heuristic model to predict flow rate," *J. Transp. Eng.*, vol. 140, no. 7, p. 04014023, 2014.
[35] D. S. Dendrinos, "Traffic-flow dynamics: A search for chaos," *Chaos, Solitons Fractals*, vol. 4, pp. 605–617, Sep. 1994.
[36] A. Stathopoulos and M. G. Karlaftis, "Spectral and cross-spectral analysis of urban traffic flows," in *Proc. IEEE Intell. Transp. Syst.*, Aug. 2001, pp. 820–825.
[37] Y. Zhang, Y. Zhang, and A. Haghani, "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 65–78, Jun. 2013.
[38] T. T. Tchrakian, B. Basu, and M. O'Mahony, "Real-time traffic flow forecasting using spectral analysis," *IEEE Transp. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 519–526, Feb. 2012.
[39] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transp. Res. C, Emerg. Technol.*, vol. 19, pp. 387–399, Sep. 2011.
[40] K. Nikola, "Evolving fuzzy neural networks-algorithms, applications and biological motivation," in *Methodologies for the Conception, Design and Application of Soft Computing*. Singapore: World Scientific, 1998, pp. 271–274.
[41] N. Kasabov, "Evolving fuzzy neural networks for supervised/unsupervised online knowledge-based learning," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 31, no. 6, pp. 902–918, Jun. 2001.
[42] N. K. Kasabov and Q. Song, "DENFIS: Dynamic evolving neural-fuzzy inference system and its application for time-series prediction," *IEEE Fuzzy Syst.*, vol. 10, no. 2, pp. 144–154, Feb. 2002.
[43] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. Upper Saddle River, NJ, USA: Prentice-Hall, 1984.
[44] T. C. Hsia, *System Identification: Least-Squares Methods*. Boston, MA, USA: D. C. Heath and Company, 1977.
[45] H. M. Adorf, "Interpolation of irregularly sampled data series—A survey," in *Proc. Astronomical Data Anal. Softw. Syst. Conf. Ser. (IVASP)*, vol. 77. 1995, pp. 460–463.
[46] T. K. Gneiting, K. Larson, M. G. Westrick, and E. Aldrich, "Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space-time method," *J. Amer. Statist. Assoc.*, vol. 101, no. 475, pp. 968–979, 2006.



Jinjun Tang received the B.S. and M.S. degrees from Shandong University of Technology, Zibo, China, in 2006 and 2009, respectively, and the Ph.D. degree in traffic information engineering and control from Harbin Institute of Technology, Harbin, China, in 2016. He was a Visiting Student with University of Washington, Seattle, WA, USA, from 2014–2016. He is currently an Associate Professor with Central South University. He has published over 30 technical papers in journals and conference proceeding as first author and co-author. His research interests include traffic flow prediction, data mining in transportation system, intelligent transportation systems, and transportation modeling.



Fang Liu received the B.S. and M.S. degrees in traffic engineering from Shandong University of Technology, Zibo, China, in 2006 and 2009, respectively. She is a Lecturer with the School of Energy and Traffic Engineering, Inner Mongolia Agricultural University, Hohhot, China. Her research interests include traffic safety, and transportation modeling and simulation.



Yajie Zou received the B.S. degree in engineering mechanics from Tongji University, Shanghai, China, and the M.S. and Ph.D. degrees in transportation engineering from Texas A&M University. He is an Associate Professor with Tongji University. His main research interests include traffic operations, traffic safety, and microscopic traffic simulation models.

Weibin Zhang received the B.S. degree in automation, the M.S. degree in mechatronics, and the Ph.D. degree in control science and engineering from Xi'an Jiaotong University, Xi'an, China. He is a Research Associate with University of Washington. His main research interests include big data analysis in transportation, traffic safety, and marine traffic.

Yinhai Wang received the Ph.D. degree from University of Tokyo, Tokyo, Japan, in 1998. He is a Professor with the Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, USA. He has published 85 peer-reviewed journal articles, made 101 invited talks, and over 170 other technical presentations. His research focuses on traffic detection systems, traffic accident modeling, transportation data management and analysis, traffic control, and traffic simulations. He is a member of the IEEE Smart Cities Steering Committee and was on the IEEE ITS Society Board of Governors from 2010 to 2013.