

## Research Article

# Short-Term Speed Prediction Using Remote Microwave Sensor Data: Machine Learning versus Statistical Model

Han Jiang,<sup>1</sup> Yajie Zou,<sup>2</sup> Shen Zhang,<sup>3</sup> Jinjun Tang,<sup>3</sup> and Yin Hai Wang<sup>4</sup>

<sup>1</sup>Department of Automation, Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing 100084, China

<sup>2</sup>Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji University, Shanghai 201804, China

<sup>3</sup>School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150001, China

<sup>4</sup>Department of Civil & Environmental Engineering, University of Washington, P.O. Box 352700, Seattle, WA 98195, USA

Correspondence should be addressed to Jinjun Tang; jinjuntang@163.com

Received 11 December 2015; Accepted 19 January 2016

Academic Editor: Valentin Lychagin

Copyright © 2016 Han Jiang 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.

Recently, a number of short-term speed prediction approaches have been developed, in which most algorithms are based on machine learning and statistical theory. This paper examined the multistep ahead prediction performance of eight different models using the 2-minute travel speed data collected from three Remote Traffic Microwave Sensors located on a southbound segment of 4th ring road in Beijing City. Specifically, we consider five machine learning methods: Back Propagation Neural Network (BPNN), nonlinear autoregressive model with exogenous inputs neural network (NARXNN), support vector machine with radial basis function as kernel function (SVM-RBF), Support Vector Machine with Linear Function (SVM-LIN), and Multilinear Regression (MLR) as candidate. Three statistical models are also selected: Autoregressive Integrated Moving Average (ARIMA), Vector Autoregression (VAR), and Space-Time (ST) model. From the prediction results, we find the following meaningful results: (1) the prediction accuracy of speed deteriorates as the prediction time steps increase for all models; (2) the BPNN, NARXNN, and SVM-RBF can clearly outperform two traditional statistical models: ARIMA and VAR; (3) the prediction performance of ANN is superior to that of SVM and MLR; (4) as time step increases, the ST model can consistently provide the lowest MAE comparing with ARIMA and VAR.

## 1. Introduction

Collecting high quality traffic information is the key factor to achieve the performance of Intelligent Transportation System (ITS). Accurate prediction of future patterns in traffic flow becomes more important in Advanced Traffic Management System (ATMS) and Advanced Traveler Information Systems (ATIS). Using the forecasted information, such as traffic volume data, travel time data, and traffic condition information, travelers can replan 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. Compared with general collecting approaches, loop detectors, and GPS equipment, Remote Traffic Microwave Sensor (RTMS) is

another important nonintrusive 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] compared 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 has 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 range from statistical prediction methods [5–10], neural networks [11–15], support vector regression [16–20], Kalman filter theory [21–26], and hybrid approaches [27–32].

In order to improve forecasting performance, the neural network model was used to aggregate speed information and acceleration information from the current forecasting segment and adjacent segments. Van Lint et al. [13] proposed a state-space neural network model that utilizes upstream and downstream traffic as model input to predict travel time with respect to missing or corrupt input data. Ma et al. [14] developed a Long Short-Term Memory Neural Network (LSTM) to predict travel speed prediction based on RTMS detection data in Beijing City; the proposed model can capture the long-term temporal dependency for time series and also automatically determine the optimal time window. For the support vector machine, Wu et al. [17] applied support vector regression (SVR) for travel time prediction, and they compared the proposed model with some traditional travel time prediction methods in highway network. Zhang and Liu [18] combined state-space approach and least squares support vector machines (LS-SVMs) to forecast travel time index. Asif et al. [20] firstly analyzed spatiotemporal trends for individual links at the network level and then constructed support vector regression (SVR) to predict travel speed in large interconnected road network. For the Kalman filter theory, Chen and Grant-Muller [22] proposed a Kalman filter type network to predict traffic flow, and they also discussed the effect of starting network parameters to the prediction performance. Chien and Kuchipudi [23] used Kalman filtering algorithm to predict travel time for its significance in continuously updating the state variable as new observations. Their empirical results indicated that the prediction performance based on historic path-based data is better than that based on link-based data during peak hours. Wang et al. [26] proposed a new extended Kalman filter (EKF) based online-learning approach to predict highway travel time. In order to effectively improve prediction performance, many scholars proposed various hybrid models to combine advantages of different kinds of methods. Dimitriou et al. [27] used genetic algorithm to structure an adaptive hybrid fuzzy rule-based system for forecasting traffic flow in urban arterial networks. Zheng et al. [28] introduced a neural network model combined with the theory of conditional probability and Bayes' rule, and the combined model that is demonstrated outperforms the singular predictors from the experimental test of Singapore's Ayer Rajah Expressway. Dong et al. [32] proposed a hybrid support vector machine that combines both statistical and heuristic models to consider the spatial-temporal patterns in traffic flow.

For the statistical model, Cetin and Comert [5] proposed a new statistical change-point detection algorithm to predict

short-term traffic flow, in which the expectation maximization and the CUSUM (cumulative sum) algorithms are implemented to detect shifts. Chandra and Al-Deek [7] considered the effect of upstream and downstream locations on the traffic at a specific location into a traditional Autoregressive Integrated Moving Average (ARIMA) model. Williams and Hoel [8] modeled a seasonal ARIMA process to complete traffic flow forecasting. For the neural networks, Ye et al. [11] used a neural network model to forecast traffic flow time series based on GPS data recorded at irregular time intervals.

According to the reviewed literature, most of traffic prediction models are mainly based on statistical methods and machine learning techniques. These two types of models have their different characteristics. The statistical models can provide good theoretical interpretability with clear calculation construction. While machine learning models use a "black box" approach to predict traffic conditions and often lack a good interpretation of the model, however, compared with statistical models, machine learning 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 [33]. In this study, we compare the prediction performances between statistical models and machine learning models. In statistical models, we select ARIMA, Vector Autoregression (VAR), and Space-Time (ST). In machine learning models, we chose Artificial Neural Network (ANN), SVM, and Multilinear Regression (MLR) as candidate. The travel speed data come from RTMS detector on fourth ring freeway in Beijing City. The contribution of this paper includes the following: comprehensively compare speed prediction performances of different models in machine learning and statistical method; analyze the prediction accuracy under different forecasting steps ahead; and evaluate models' performance under different scenarios.

The remainder of paper is organized as follows. Section 2 briefly introduces the models used in this study. The data source and analysis are provided in Section 3. Section 4 discusses the results and compares prediction accuracies of different models. Section 5 provides the conclusion of the paper.

## 2. Methodology

**2.1. Statistical Models.** In this section, we briefly introduce three statistical methods (i.e., ARIMA, VAR, and ST) considered in this study.

**2.1.1. ARIMA.** The Autoregressive Integrated Moving Average (ARIMA)  $(p, d, q)$  model contains the following parameters:  $p$  is the number of autoregressive terms,  $d$  is the number of nonseasonal differences, and  $q$  is the number of lagged forecast errors. An ARIMA model can be regarded as a generalization of autoregressive moving average (ARMA) model. The mathematical formulation of an ARMA  $(p, q)$  process is defined as follows:

$$\begin{aligned} x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \cdots - \phi_p x_{t-p} \\ = \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \cdots + \theta_q \omega_{t-q}, \end{aligned} \quad (1)$$

where  $\{x_t\}$  is stationary,  $\omega_t$  is a normal white noise series with mean zero and variance  $\sigma_w^2$ ,  $\{\phi_1, \dots, \phi_p\}$  and  $\{\theta_1, \dots, \theta_q\}$  are parameters for the autoregressive and the moving average terms, and the polynomials  $(1 - \phi_1 z - \dots - \phi_p z^p)$  and  $(1 + \theta_1 z + \dots + \theta_q z^q)$  have no common factors. Assuming  $B^k x_t = x_{t-k}$ ,  $B^k z_t = \omega_{t-k}$ ,  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ , and  $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$ , the ARMA  $(p, q)$  model can be written as follows:

$$\phi(B) X_t = \theta(B) Z_t. \quad (2)$$

The ARMA model requires that the data series are stationary. When time series data are nonstationary, the ARIMA model is proposed to model the data which does not show evidence of an ARMA model. In the ARIMA model, the integrated part with order  $d$ , denoted as  $I(d)$ , means the  $d$ th difference of the original data, which can transform the original data to a stationary series. The mathematical equation of an ARIMA  $(p, d, q)$  model is

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) (1 - B)^d X_t \\ = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) Z_t. \end{aligned} \quad (3)$$

**2.1.2. VAR.** The Vector Autoregression (VAR) model can capture the linear interdependencies among multiple time series and thus can consider the effect of the neighboring stations in predicting the future speed. Here, a 3-equation VAR( $m$ ) model is used and its formulation is defined as follows:

$$\mathbf{X}_{t+1} = \boldsymbol{\varphi}_0 + \boldsymbol{\varphi}_1 \mathbf{X}_t + \boldsymbol{\varphi}_2 \mathbf{X}_{t-1} + \dots + \boldsymbol{\varphi}_m \mathbf{X}_{t-m+1} + \mathbf{u}_{t+1}, \quad (4)$$

where  $\mathbf{X}_{t+1} = (A_{t+1}, B_{t+1}, C_{t+1})^T$  is the  $3 \times 1$  vector of variables,  $\boldsymbol{\varphi}_0$  is the  $3 \times 1$  constant term,  $\boldsymbol{\varphi}_1$  through  $\boldsymbol{\varphi}_m$  are  $3 \times 3$  coefficient matrices, and  $\mathbf{u}_{t+1}$  is the corresponding  $3 \times 1$  independently and identically distributed random vector with  $E(\mathbf{u}_{t+1}) = 0$  and time invariant positive definite covariance matrix  $E(\mathbf{u}_{t+1} \mathbf{u}_{t+1}^T) = \Sigma_u$ . Before applying the VAR( $m$ ) model, the characteristic polynomial is evaluated to ensure the stability:

$$\det(I_3 - \boldsymbol{\varphi}_1 z - \dots - \boldsymbol{\varphi}_m z^m) \neq 0 \quad \text{for } |z| \leq 1, \quad (5)$$

where  $I_3$  is a  $3 \times 3$  identity matrix. The necessary and sufficient condition for stability is that all characteristic roots lie outside the unit circle.

**2.1.3. Space-Time Model.** The Space-Time (ST) model is a probabilistic modeling approach that can provide the point prediction of future observations [10]. In probabilistic speed prediction, the commonly used normal distribution is adopted for speed data. Thus, this study assumed that the speed at time  $t + p$  at the target station,  $S_{t+p} \sim N(\mu_{t+p}, \sigma_{t+p}^2)$ , follows a normal distribution. The point prediction of  $S_{t+p}$  is the mean,  $\mu_{t+p}$ , of the normal distribution. Then,  $\mu_{t+p}$  is fitted by a linear combination of the present and past values of the

speed series at all stations. For example, for station B, when  $p = 1$  (i.e., 2-minute ahead prediction),

$$\begin{aligned} \mu_{t+1} = \alpha_0 + \alpha_1 S_{B,t} + \alpha_2 S_{B,t-1} + \alpha_3 S_{A,t} + \alpha_4 S_{A,t-1} \\ + \alpha_5 S_{C,t-2}, \end{aligned} \quad (6)$$

where  $S_{A,t}$ ,  $S_{B,t}$ , and  $S_{C,t}$  are the 2-minute average speed at stations A, B, and C at time  $t$ ; stations A and C are the upstream and downstream of station B, and  $\alpha_0, \alpha_1, \dots, \alpha_5$  are model coefficients. Predictor variables for  $\mu_{t+p}$  are selected based on an analysis of the speed data from first week of the dataset using a stepwise forward search (refer to [10] for details about predictor variable selection algorithm).

**2.2. Machine Learning Models.** In this section, we select three models, Artificial Neural Network, support vector machine, and Multilinear Regression, to predict travel speed; the following subsections briefly describe these three models.

**2.2.1. Artificial Neural Network.** Artificial Neural Network (ANN) is a popular tool for traffic flow prediction because of its capability of handling multidimensional data, flexible model structure, strong generalization and learning ability, and adaptability [33]. Different from the statistical methods, ANN does not require underlying assumptions regarding data and is also robust to missing and noisy inputs [33]. ANN model is generally constructed as multilayer system and it is typically defined by three types of parameters: the interconnection pattern between different layers; the learning process for updating the weights for the layers; the activation function that converts input to output activation. An ANN system can be represented as follows:

$$y_{\text{ann}} = h \left( \varphi_0 + \sum_{j=1}^N \varphi_j g \left( \theta_0 + \sum_{i=1}^M \theta_i x_i \right) \right), \quad (7)$$

where  $M$  and  $N$ , respectively, represent the number of neurons in the input layer and hidden layer and  $g$  and  $h$  are the transfer functions for the input layer and hidden layer. The vector matrices of  $\mathbf{h}$  and  $\mathbf{u}$ , respectively, refer to the weight values for neurons in both input layer and hidden layer. To minimize the sum of estimated errors from ANN, a number of optimization algorithms were developed including Back Propagation Neural Networks, Levenberg-Marquardt method, and genetic algorithm. The detailed information about ANN is introduced in [11–15, 33]. As an important member in ANN family, nonlinear autoregressive model with exogenous inputs neural network (NARXNN) allows a delay line on the inputs, and the outputs feed back to the input by another delay line. This is a further extension of the time delay neural network since the NARXNN not only considers its own previous outputs but also incorporates the exogenous inputs [14].

**2.2.2. Support Vector Machine.** The main idea of support vector regression is to map data into a high-dimensional feature space through a nonlinear relationship and then construct a linear regression in this 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. The SVM regression function is formulated as follows:

$$f(x) = \mathbf{w}^T \Phi(x) + b, \quad (8)$$

where  $\mathbf{w}$  is a vector in a feature space  $F$  and  $\Phi(x)$  is called the feature, which maps the input  $x$  to a vector in  $F$ . Assume an  $\varepsilon$ -insensitive loss function:

$$L_\varepsilon(x, y, f) = |y - f(x)|_\varepsilon = \max(0, |y - f(x)| - \varepsilon). \quad (9)$$

Then,  $w$  and  $b$  are estimated by solving the following optimization problem:

$$C \sum_{i=1}^N L_\varepsilon(x_i, y_i, f) + \frac{1}{2} \|\mathbf{w}\|^2, \quad (10)$$

where  $\varepsilon$  is the maximum deviation allowed;  $C$  represents the associated penalty for expressing deviation during the training process, which evaluates the trade-off between the empirical risk and the smoothness of the model. The positive slack variables  $\beta$  and  $\beta^*$  are incorporated, which represent the size of positive and negative excess deviation, respectively. Thus, (10) is transformed to the following constrained formulation:

$$\begin{aligned} \min \quad & J(\beta_i, \beta_i^*) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N (\beta_i + \beta_i^*) \\ \text{subject to:} \quad & \mathbf{w}^T \Phi(x) + b - y_i \leq \varepsilon + \beta_i^* \\ & y_i - \mathbf{w}^T \Phi(x) - b \leq \varepsilon + \beta_i \\ & \beta_i, \beta_i^* \geq 0. \end{aligned} \quad (11)$$

The first term of (11),  $\|\mathbf{w}\|^2$ , is the regularized term. Thus, it controls the function capacity. The second term,  $\sum_{i=1}^N (\beta_i + \beta_i^*)$ , is the empirical error measured by  $\varepsilon$ -insensitive loss function. By using the appropriate Karush-Kuhn-Tucker (KKT) conditions to (11), we have the following dual form of the optimization problem:

$$\begin{aligned} \max \quad & Q(\alpha_i, \alpha_i^*) \\ & = \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^N (\alpha_i - \alpha_i^*) \\ & \quad - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i - x_j) \end{aligned} \quad (12)$$

$$\begin{aligned} \text{subject to:} \quad & \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ & 0 \leq \alpha_i \leq C \\ & 0 \leq \alpha_i^* \leq C. \end{aligned}$$

Therefore, the SVM equation for nonlinear predictions becomes

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x, x_i) + b, \quad (13)$$

where  $K(x, x_i)$  is called the kernel function.  $\alpha_i$  and  $\alpha_i^*$  are the solution to the dual problem. There are four conventional kernel functions: linear, radial basis function (RBF), polynomial, and sigmoid. In this study, we select two common functions, linear and RBF, to construct SVM model. The first reason is that these two functions are widely used in prediction and classification. For the second reason, the linear function has advantages of simple construction and low computational time; the RBF function uses nonlinear structure and produces reliable prediction performance based on optimal parameters. For the linear function,

$$K(x_i, x_j) = x_i^T x_j. \quad (14)$$

For the RBF kernel functions,

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right), \quad (15)$$

where  $\gamma$  is the parameter in kernel function.

**2.2.3. Multilinear Regression.** 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:

$$y_{\text{mlr}}(k) = a_0 + \sum_{i=1}^l a_i \cdot y_i(k-l). \quad (16)$$

$y_{\text{mlr}}(k)$  represents the prediction value at the  $k$ th period. The independent variable  $y_i(k-l)$  means the speed data at the previous  $(k-l)$ th period,  $l$  is the number of historical data considered in MLR, and  $a_0$  and  $a_i$  are the regression parameters which can be optimized by training samples. The prediction values in testing dataset are estimated from (16).

### 3. Data Description

The travel speed data used in the study were collected in 4th ring road in Beijing. The segment we 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 Figure 1. The distance between each two adjacent stations is about 1.4 km. Location A represents detector 9052, location B is detector 9053, and location C indicates detector 9054. 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 that the selected speed data are appropriate for model validation and evaluation in this study. The data collection duration starts from December 1, 2014, to December 31, 2014, total of 31 days. In order to validate the prediction performance of different models and fairly compare the prediction accuracy, we divide data 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

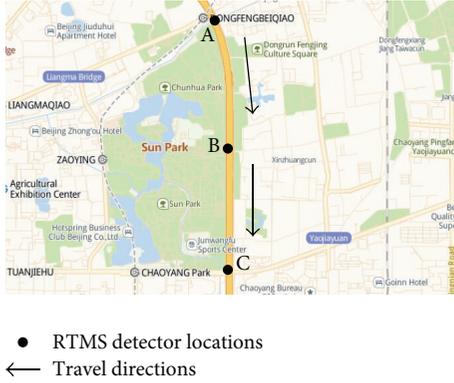


FIGURE 1: Three data collection stations in major ring road, Beijing.

last 10 days are employed to validate models effectiveness. The data in first 7 days from each station are plotted in Figure 2 to show the general trends. For three stations, we can see that speed data distribute similar patterns but different speed values. Figure 2 demonstrates clearly sharp reduction of speed during peak hours and also shows that traffic during the night is normally smooth without fluctuation. The speed data collected in locations A and B have similar distribution patterns, and they express obvious periodicity with low speed at evening peak hours. The speed detected in location C represents different patterns compared with A and B. The speed values are lower in evening peak hours than other locations, because traffic here is under high pressure and volumes are much higher in evening peak hours. The limitation of data includes erroneous samples and data missing. To the inaccurate data, for example, speed values are higher than speed limit or speed values are negative, we remove those samples from the original dataset. Furthermore, the data missing can be attributed to many natural and man-made reasons, for example, communication failures, malfunctioning devices, incorrect observations, or data transfer problems. So, aimed at the data collection shortcoming, historical averaged based method has been implemented to impute missing and removed data, which ensures that the selected speed samples are appropriate for model validation and evaluation in this study.

#### 4. Models Comparison and Results Discussion

In ANN, Back Propagation Neural Network (BPNN) and NARXNN are selected as the candidate models in comparison, and they both have one hidden layer with 50 neurons. We use neural network tool in MATLAB to optimize parameters. In SVM, we use RBF and linear structure as kernel functions. For the parameters optimization, [31] 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*. The coefficients of the ST model are estimated using the minimum continuous ranked probability score (CRPS) estimation [34]. When forecasting future speed values, the best order of the ARIMA model is determined by the 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.

For all the prediction algorithms, the data in first 21 days are used for training models and data in last 10 days are used for validating models. Two performance measures including the mean absolute error (MAE) and the mean absolute percentage error (MAPE) are used as indicators to evaluate the multistep prediction performances of different models. The unit of the MAE is km per hour. The equations for calculating MAE and MAPE are shown as follows:

$$\text{MAE} = \frac{\sum_{i=1}^N |\bar{v}_i - v_i|}{N}, \quad (17)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|\bar{v}_i - v_i|}{v_i} \times 100\%,$$

where  $N$  is the number of observations,  $v_i$  is the actual speed at time  $i$  at station, and  $\bar{v}_i$  is the predicted speed. Furthermore, in order to further evaluate the performance of all models, both one-step and multistep ahead prediction (i.e., 3-step (6 minutes), 5-step (10 minutes), and 10-step (20 minutes)) are considered. Tables 1, 2, and 3 provide the MAE and MAPE values of different models for different forecasting horizons. Note that, in Table 1, bold values indicate the smallest MAE and MAPE values. Figure 3 shows the prediction results of models for one forecasting step compared with observed speed data in five days. The left column represents the prediction results of machine learning methods and right column represents the prediction results of statistical models. In machine learning, we only show comparison among BPNN, SVM-RBF, and MLR. In statistical models, only prediction performances of ARIMA and ST are compared. Figure 4 displays the correlation between observed values and predicted values from five models (three models in machine learning and two models in statistical method) in ten days, and  $R^2$  represents the correlation coefficient to evaluate the relevance between observed and predicted values. Figure 5 shows the frequency distribution of predicted errors of five models in ten days. The  $x$ -axis is the errors (errors = predicted value – observed value), and  $y$ -axis indicates the frequency in different error ranges. In the figure,  $r$  is defined as the rate percentage that errors fell within the range of  $\pm 10\%$  to estimate prediction strength of all models. Figure 6 represents the frequency distribution of relative errors of five models in ten days. The  $x$ -axis is the relative errors (relative errors = (predicted value – observed value)/observed value), and  $y$ -axis indicates the frequency in different error ranges. Similarly, in Figure 6,  $r$  is defined as the rate percentage that relative errors fell within the range of  $\pm 10\%$  to estimate the performance of models.

From the observation of four figures, we can gain several conclusions as follows: the BPNN produce better prediction results with higher  $R^2$  and  $r$  compared with other models; SVM-RBF outperforms ARIMA. The reason is that machine learning models have complex structure and strong learning ability; considering correlation between spatial and temporal

TABLE 1: Prediction accuracy of models for different forecasting steps ahead in station A.

MAE	Number of forecasting steps ahead			
	1	3	5	10
BPNN	<b>2.5027</b>	3.0243	3.4576	4.3553
NARXNN	2.5041	<b>3.0232</b>	<b>3.4482</b>	4.3357
SVM-RBF	2.6537	3.1135	3.4858	4.2625
SVM-LIN	2.7492	3.1160	3.4714	4.2604
MLR	3.0820	3.4795	3.8774	4.6338
ARIMA	2.6777	3.1501	3.5291	4.3233
VAR	2.6835	3.2931	3.7145	4.6030
ST	2.9398	3.1414	3.4932	<b>4.2355</b>

MAPE (%)	Number of forecasting steps ahead			
	1	3	5	10
BPNN	<b>5.2831</b>	<b>6.4927</b>	7.5472	9.5735
NARXNN	5.2923	6.4995	<b>7.4855</b>	9.6267
SVM-RBF	5.5839	6.6952	7.5901	<b>9.4775</b>
SVM-LIN	5.2878	6.5408	7.6922	9.4937
MLR	5.9048	7.3508	8.1869	9.9782
ARIMA	5.6290	6.7794	7.6974	9.6376
VAR	5.6271	6.7202	8.2556	9.5937
ST	6.1660	6.7725	7.6390	9.5080

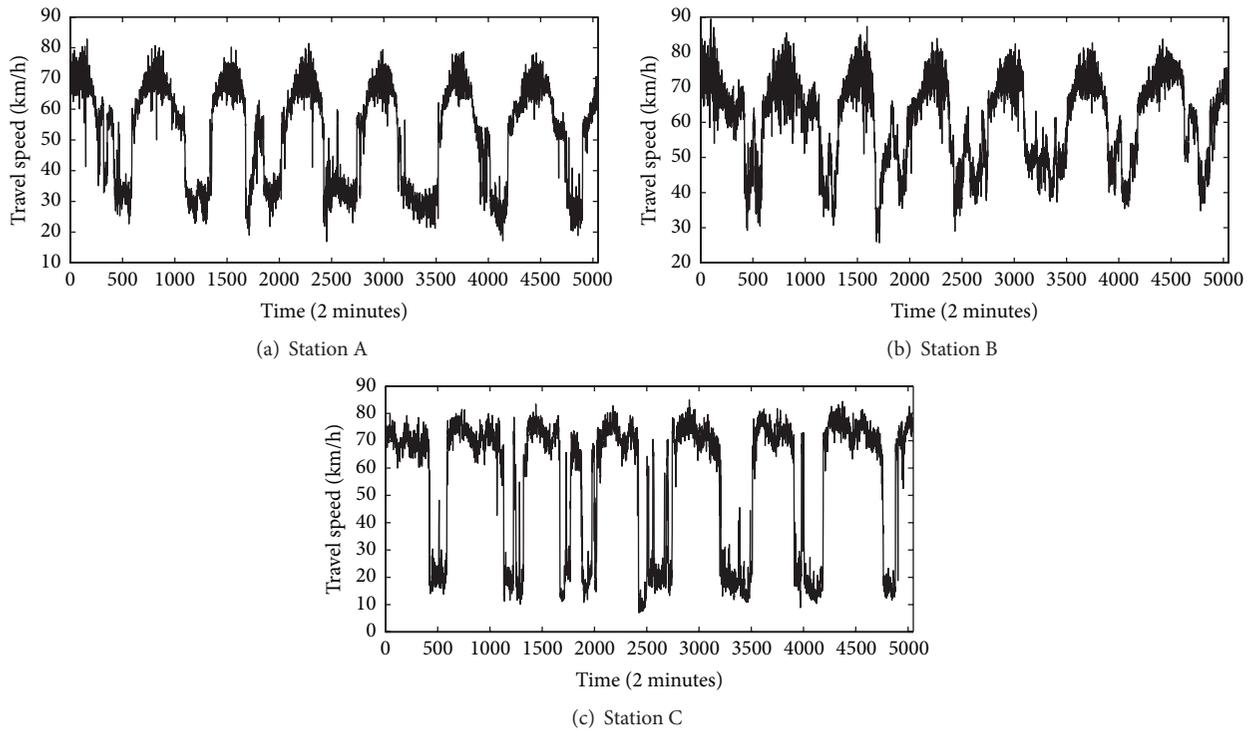


FIGURE 2: Travel speed distribution of three collection locations in 7 days.

characteristics, ST also produces high prediction accuracy; it has similar prediction accuracy with SVM-RBF; MLR is inferior to ARIMA and SVM-RBF with lowest  $R^2$  and  $r$  for its simple calculation structure.

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 1, 2, and 3 show that the MAE and MAPE values for 10-step ahead forecasting are significantly larger than the results of 1-step ahead. From the observation of figures, we

TABLE 2: Prediction accuracy of models for different forecasting steps ahead in station B.

MAE	Number of forecasting steps ahead			
	1	3	5	10
BPNN	2.9402	3.1513	3.3712	<b>3.8135</b>
NARXNN	<b>2.9297</b>	<b>3.1434</b>	<b>3.3591</b>	3.8208
SVM-RBF	2.9825	3.2323	3.4342	3.8567
SVM-LIN	3.1945	3.2111	3.5712	3.8806
MLR	3.2624	3.9019	4.1148	4.5207
ARIMA	3.0144	3.2669	3.4683	3.8959
VAR	3.0244	3.3033	3.7154	4.0356
ST	3.4509	3.2993	3.5093	3.9322
MAPE (%)	Number of forecasting steps ahead			
	1	3	5	10
BPNN	4.9518	5.3793	5.7737	<b>6.6007</b>
NARXNN	<b>4.9325</b>	<b>5.3470</b>	<b>5.7459</b>	6.6096
SVM-RBF	4.9811	5.4445	5.8253	6.6268
SVM-LIN	5.4039	5.4243	6.2740	6.8330
MLR	5.5909	6.4957	6.8970	7.6689
ARIMA	5.0383	5.5078	5.8899	6.7010
VAR	5.0470	5.6490	6.4091	6.8736
ST	5.7732	5.5666	5.9707	6.7795

TABLE 3: Prediction accuracy of models for different forecasting steps ahead in station C.

MAE	Number of forecasting steps ahead			
	1	3	5	10
BPNN	2.5694	<b>3.4613</b>	4.1736	5.6407
NARXNN	<b>2.5484</b>	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
MLR	3.9650	3.8713	4.5163	5.9062
ARIMA	2.6675	3.6235	4.2857	5.6745
VAR	2.6611	3.6798	4.3051	5.4380
ST	2.6359	3.5298	<b>4.0963</b>	<b>5.2024</b>
MAPE (%)	Number of forecasting steps ahead			
	1	3	5	10
BPNN	6.5606	<b>8.9994</b>	11.3564	16.3149
NARXNN	<b>6.4875</b>	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
MLR	7.1717	10.4244	11.3601	16.7207
ARIMA	6.7822	9.4230	11.6734	16.7901
VAR	6.6898	9.1819	11.0211	14.5708
ST	6.6709	<b>8.9993</b>	<b>10.7423</b>	<b>14.3994</b>

can obtain similar results: as the step of prediction increases, the predicted values of all models become more fluctuated, and the prediction accuracy and stability of multistep ahead decrease compared to the results of 1-step ahead prediction.

- (2) 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. Considering correlation between spatial and temporal, ST also produces high prediction accuracy. SVM-LIN produces similar prediction results with ARIMA, and MLR is inferior to ARIMA and VAR.

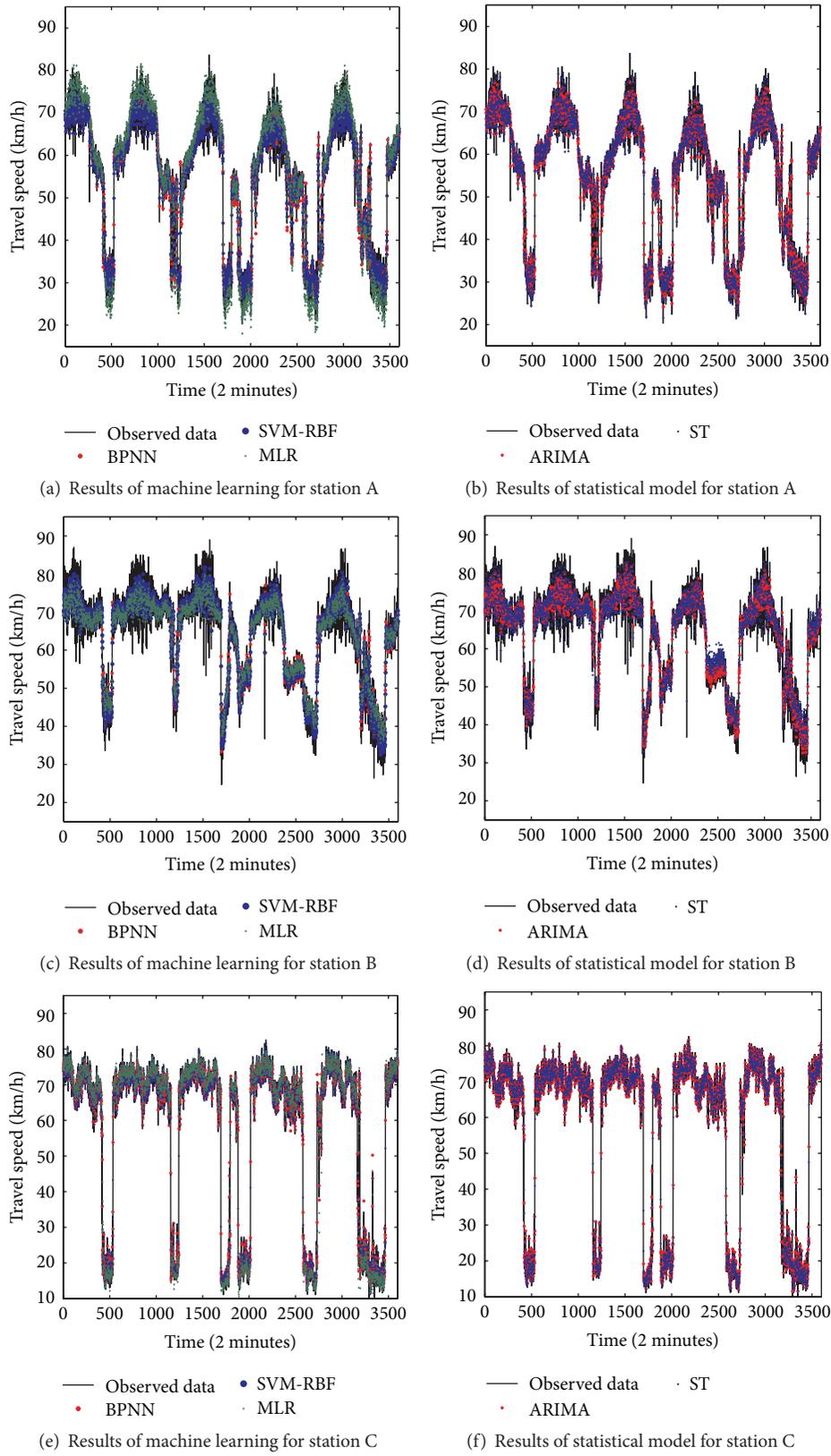


FIGURE 3: Prediction result for one step ahead in three stations.

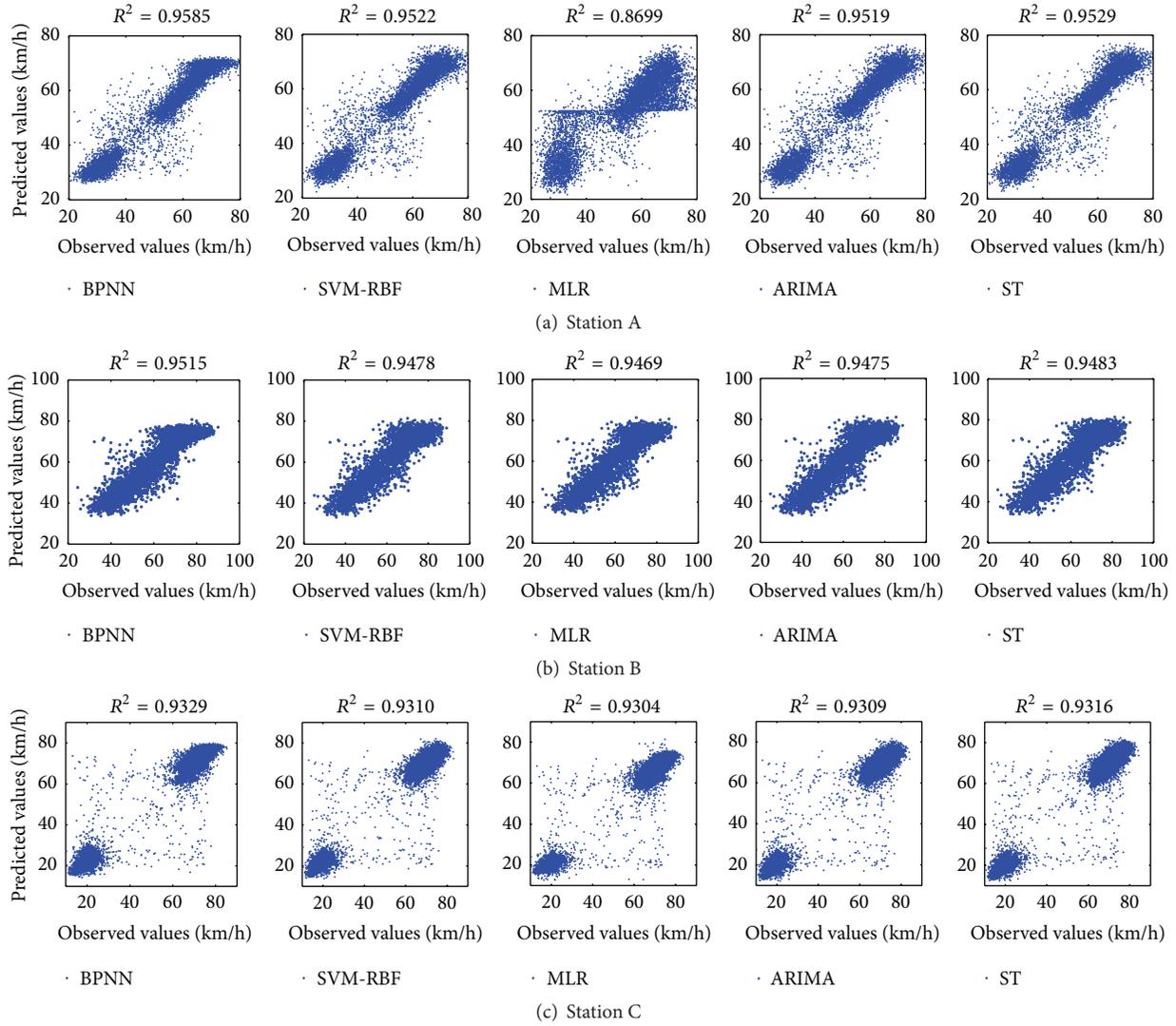


FIGURE 4: Correlation distribution of predicted result for three steps ahead in three stations.

- (3) For the models in machine learning, the prediction performance of ANN is superior to that of SVM and MLR. Furthermore, BP and NARX both achieve similar high accuracy. In the SVM, as RBF is more flexible than linear function, the SVM-RBF outperforms SVM-LIN clearly. As for simple calculation structure, the prediction accuracy of MLR is lower than that of NN and SVM.
- (4) For the statistical models, as the time step increases, the ST model can consistently provide the lowest MAE values comparing with ARIMA and VAR. It can be observed that the ST model is preferred over the VAR and ARIMA models when predicting multiple time points into the future. The possible reason is that the ST model uses spatial and temporal information from neighboring stations observed at time  $t$  to directly predict the future speed value at time  $t+p$ . On the contrary, the ARIMA and VAR models consider  $t+1$ 's prediction as an observed value and use it

to predict speed at time  $t+2$  and this procedure is repeated  $p$  times to forecast speed values at time  $t+p$ . Thus, the prediction error may accumulate after multiple steps when using ARIMA and VAR models.

- (5) For three stations, we observed that the predicted accuracy in stations A and B is relatively higher than that in station C from Tables 1, 2, and 3 and Figures 3, 4, 5, and 6. The reason is that the speed data collected in station C vary sharply from low values to high values, which can be seen in Figure 2. With different distribution patterns in stations A and B, the speed increases rapidly from 20 km/h to about 80 km/h in station C. Moreover, from the observation from Figure 4, we can see that the speed samples in station C are mainly classified into two clusters: low value part and high value part. Due to the lack of the part with continuous increase of speed values, the sharp fluctuation in speed distribution will definitely influence predicted performance of models.

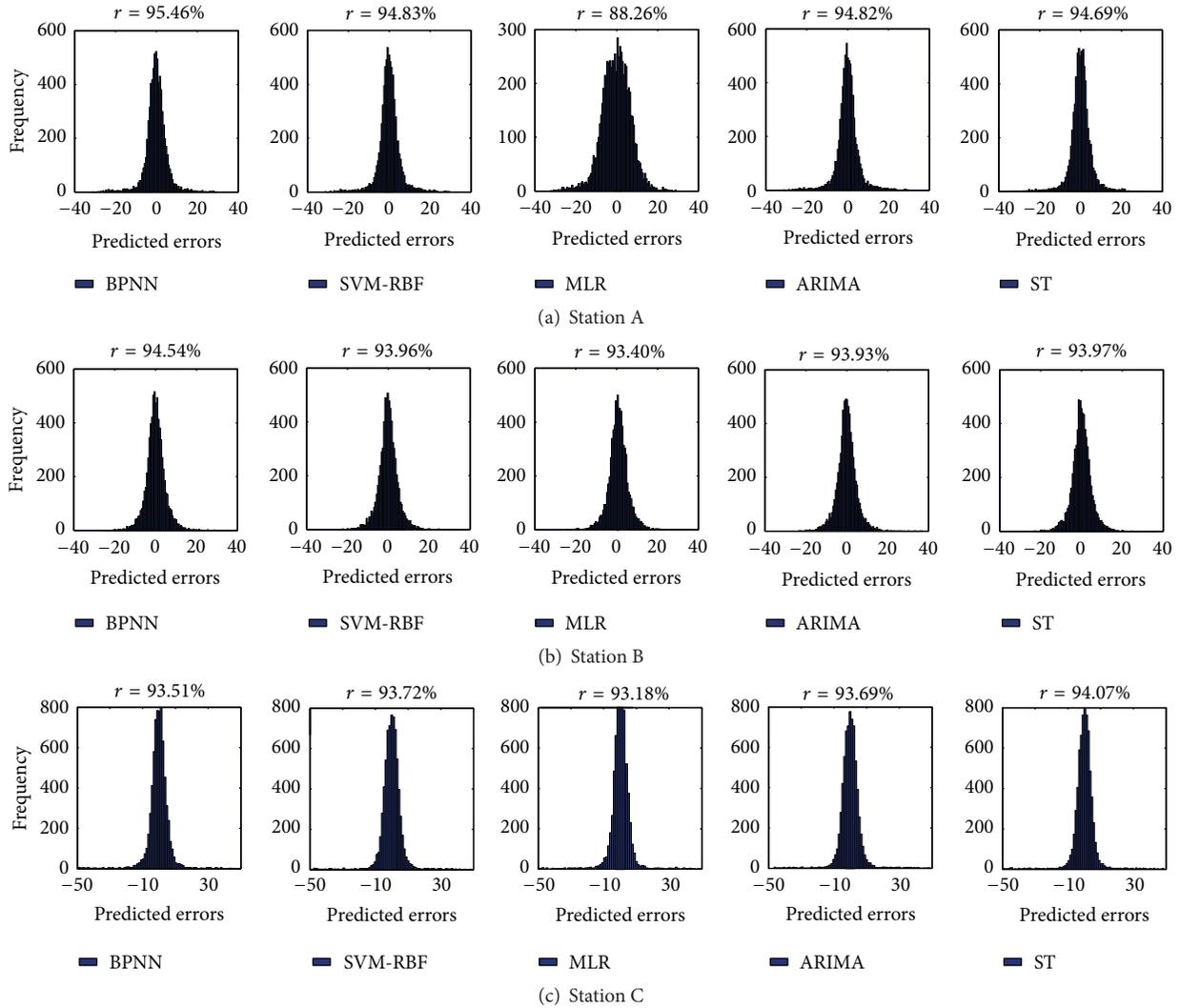


FIGURE 5: Frequency distribution of predicted errors for five steps ahead in three stations.

Although from the prediction results above we can find that complex machine learning models achieve higher prediction results than statistical methods, it is still a challenging work to select proper model in actual applications. As Karlaftis and Vlahogianni [33] suggested, prediction accuracy is a very important indicator but model simplicity and suitability also should be considered. Kirby et al. [35] stated that accuracy should not be the sole determinant for selecting the proper methodology for prediction; other issues should be considered in selecting the appropriate approach such as the time and effort required for model development, skills and expertise required, transferability of the results, and adaptability to changing behaviors [33, 35]. Between classical statistical models and machine learning algorithms, researchers frequently prefer higher prediction accuracy over explanatory power of model. Karlaftis and Vlahogianni [33] concluded some criteria for model selection between NN and statistical models. Thus, we applied case study to compare the predicted performance between machine learning and statistical models, and the conclusion we obtained in this study

is that complex machine learning models have advantages of nonlinear fitting ability and robustness for missing data but low explanatory power, and statistical methods can reach high prediction performance through inherent explanatory structure.

### 5. Conclusions

This paper evaluated the multistep prediction performance of machine learning and statistical models using the speed data collected from three RTMS located on 4th ring road in Beijing City. The data are collected from December 1, 2014, to December 31, 2014, with interval of 2 minutes. In the models performance comparison, we choose five machine learning methods: BPNN, NARXNN, SVM-RBF, SVM-LIN, and MLR, and three conventional statistic models: ARIMA, VAR, and ST model. We firstly provide a brief introduction of each model. In the applications, we then optimize models parameters by using data collected in the first 21 days. Finally, we compare prediction accuracies of different models based

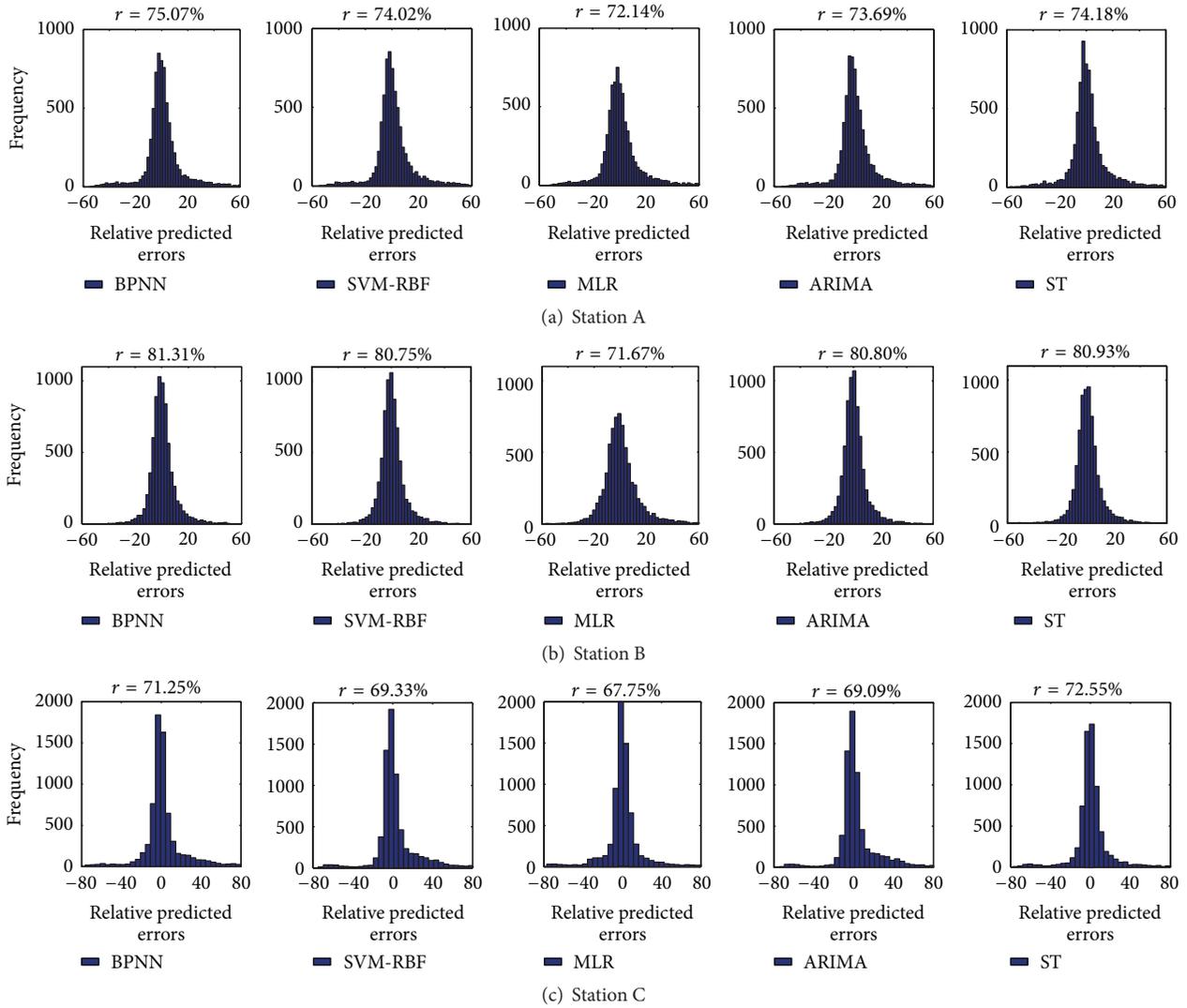


FIGURE 6: Frequency distribution of relative errors for ten steps ahead in three stations.

on data collected in last 10 days. Based on observation of the results from this study, several interesting conclusions can be drawn. First, the prediction accuracy of speed decreases as the prediction time steps increase for all models. Second, the BPNN, NARXNN, and SVM-RBF are superior to two traditional statistical models: ARIMA and VAR. Third, among the machine learning models, the prediction performance of ANN is better than that of SVM and MLR. Fourth, in the statistical models, as the time step increases, the ST model can consistently provide the lowest MAE values comparing with ARIMA and VAR. The statistical models have good theoretical interpretability and machine learning models are more flexible with no or little prior assumptions for input variables. In addition, machine learning models are more capable of processing outliers, missing and noisy data. Through case study in Beijing based on observed data, this study supplies useful applications of speed prediction and comparisons of prediction performance.

**Conflict of Interests**

The authors declare that there is no conflict of interests regarding the publication of this paper.

**Acknowledgments**

This research was partly supported by the National Natural Science Foundation of China (Grant nos. 51138003, 51329801, and G030601), Natural Science Foundation of Heilongjiang Province (Grant no. QC2013C047), and Hi-Tech Research and Development Program of China (Grant no. 2014BAG03B04).

**References**

[1] X. Yu and P. D. Prevedouros, “Performance and challenges in utilizing non-intrusive sensors for traffic data collection,” *Advances in Remote Sensing*, 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," *Transport Reviews*, 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," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 3–19, 2014.
- [4] J. W. C. Van Lint and C. P. I. J. Van Hinsbergen, "Short term traffic and travel time prediction models, in artificial intelligence applications to critical transportation issues," in *Transportation Research Circular*, R. Chowdhury and S. Sadek, Eds., Number E-C168, National Academies Press, Washington, DC, USA, 2012.
- [5] M. Cetin and G. Comert, "Short-term traffic flow prediction with regime switching models," *Transportation Research Record*, no. 1965, pp. 23–31, 2006.
- [6] P. Shang, X. Li, and S. Kamae, "Chaotic analysis of traffic time series," *Chaos, Solitons and Fractals*, vol. 25, no. 1, pp. 121–128, 2005.
- [7] S. R. Chandra and H. Al-Deek, "Cross-correlation analysis and multivariate prediction of spatial time series of freeway traffic speeds," *Transportation Research Record*, no. 2061, pp. 64–76, 2008.
- [8] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 664–672, 2003.
- [9] 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.
- [10] 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.
- [11] Q. Ye, W. Y. Szeto, and S. C. Wong, "Short-term traffic speed forecasting based on data recorded at irregular intervals," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1727–1737, 2012.
- [12] K. Y. Chan, T. S. Dillon, J. Singh, and E. Chang, "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg-Marquardt algorithm," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 644–654, 2012.
- [13] 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," *Transportation Research Part C*, vol. 13, no. 5–6, pp. 347–369, 2005.
- [14] 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," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187–197, 2015.
- [15] L. R. Rilett and D. Park, "Direct forecasting of freeway corridor travel times using spectral basis neural networks," *Transportation Research Record*, no. 1752, pp. 140–147, 2001.
- [16] S. H. Lam and T. D. Toan, "Short-term travel time predictions using support vector regression," in *Proceedings of the Transportation Research Board Annual Meeting*, CD-ROM, National Academies Press, Washington, DC, USA, January 2008.
- [17] C.-H. Wu, J.-M. Ho, and D. T. Lee, "Travel-time prediction with support vector regression," *IEEE Transactions on Intelligent Transportation Systems*, vol. 5, no. 4, pp. 276–281, 2004.
- [18] Y. Zhang and Y. Liu, "Traffic forecasting using least squares support vector machines," *Transportmetrica*, vol. 5, no. 3, pp. 193–213, 2009.
- [19] Y. Zhang and Y. Xie, "Forecasting of short-term freeway volume with v-support vector machines," *Transportation Research Record*, no. 2024, pp. 92–99, 2007.
- [20] M. T. Asif, J. Dauwels, C. Y. Goh et al., "Spatiotemporal patterns in large-scale traffic speed prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 794–804, 2014.
- [21] Y.-C. Zhao, Y.-C. Liu, L.-Q. Shan, and B. Zhou, "Dynamic analysis of Kalman filter for traffic flow forecasting in sensor networks," *Information Technology Journal*, vol. 11, no. 10, pp. 1508–1512, 2012.
- [22] H. Chen and S. Grant-Muller, "Use of sequential learning for short-term traffic flow forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 9, no. 5, pp. 319–336, 2001.
- [23] S. I.-J. Chien and C. M. Kuchipudi, "Dynamic travel time prediction with real-time and historic data," *Journal of Transportation Engineering*, vol. 129, no. 6, pp. 608–616, 2003.
- [24] S. I. J. Chien, X. Liu, and K. Ozbay, "Predicting travel times for the South Jersey real-time motorist information system," *Transportation Research Record*, vol. 1855, pp. 32–40, 2003.
- [25] J. W. C. Van Lint, "Online learning solutions for freeway travel time prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 1, pp. 38–47, 2008.
- [26] Y. Wang, M. Papageorgiou, and A. Messmer, "RENAISSANCE: a unified macroscopic model-based approach to real-time freeway network traffic surveillance," *Transportation Research Part C: Emerging Technologies*, vol. 14, no. 3, pp. 190–212, 2006.
- [27] L. Dimitriou, T. Tsekeris, and A. Stathopoulos, "Adaptive hybrid fuzzy rule-based system approach for modeling and predicting urban traffic flow," *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 5, pp. 554–573, 2008.
- [28] W. Zheng, D.-H. Lee, and Q. Shi, "Short-term freeway traffic flow prediction: bayesian combined neural network approach," *Journal of Transportation Engineering*, vol. 132, no. 2, pp. 114–121, 2006.
- [29] K. Hamad, M. T. Shourijeh, E. Lee, and A. Faghri, "Near-term travel speed prediction utilizing Hilbert-Huang transform," *Computer-Aided Civil and Infrastructure Engineering*, vol. 24, no. 8, pp. 551–576, 2009.
- [30] Y. Xie, Y. Zhang, and Z. Ye, "Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition," *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 326–334, 2007.
- [31] 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, pp. 21–31, 2014.
- [32] C. Dong, S. H. Richards, Q. Yang, and C. Shao, "Combining the statistical model and heuristic model to predict flow rate," *Journal of Transportation Engineering*, vol. 140, no. 7, 2014.
- [33] M. G. Karlaftis and E. Vlahogianni, "Statistical methods versus neural networks in transportation research: differences, similarities and some insights," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 3, pp. 387–399, 2011.

- [34] T. Gneiting, K. Larson, K. Westrick, M. G. Genton, and E. Aldrich, "Calibrated probabilistic forecasting at the stateline wind energy center: the regime-switching space-time method," *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 968–979, 2006.
- [35] H. R. Kirby, S. M. Watson, and M. S. Dougherty, "Should we use neural networks or statistical models for short-term motorway traffic forecasting?" *International Journal of Forecasting*, vol. 13, no. 1, pp. 43–50, 1997.



# Hindawi

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

