JOURNAL OF ADVANCED TRANSPORTATION
J. Adv. Transp. 2016; **50**:489–506
Published online 17 December 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/atr.1356

Urban traffic flow prediction: a spatio-temporal variable selection-based approach

Yanyan Xu[1], Hui Chen[2]*, Qing-Jie Kong[3], Xi Zhai[4] and Yuncai Liu[1]

[1]*Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China*
[2]*School of Information Science and Engineering, Shandong University, Jinan 250100, China*
[3]*Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*
[4]*Shanghai Transportation Information Center, Shanghai Urban-Rural Construction and Transportation Development Research Institute, Shanghai 200032, China*

SUMMARY

Short-term traffic flow prediction in urban area remains a difficult yet important problem in intelligent transportation systems. Current spatio-temporal-based urban traffic flow prediction techniques trend aims to discover the relationship between adjacent upstream and downstream road segments using specific models, while in this paper, we advocate to exploit the spatial and temporal information from all available road segments in a partial road network. However, the available traffic states can be high dimensional for high-density road networks. Therefore, we propose a spatio-temporal variable selection-based support vector regression (VS-SVR) model fed with the high-dimensional traffic data collected from all available road segments. Our prediction model can be presented as a two-stage framework. In the first stage, we employ the multivariate adaptive regression splines model to select a set of predictors most related to the target one from the high-dimensional spatio-temporal variables, and different weights are assigned to the selected predictors. In the second stage, the kernel learning method, support vector regression, is trained on the weighted variables. The experimental results on the real-world traffic volume collected from a sub-area of Shanghai, China, demonstrate that the proposed spatio-temporal VS-SVR model outperforms the state-of-the-art. Copyright © 2015 John Wiley & Sons, Ltd.

KEY WORDS: traffic flow prediction; urban road network; spatio-temporal correlation; high-dimensional regression; variable selection

1. INTRODUCTION

With the frequent occurrence of traffic congestion in the urban road network, reliable short-term traffic flow prediction is becoming an extremely crucial task in intelligent transportation systems (ITS), especially in the advanced traveler information systems and the advanced transportation management systems. Accurate traffic state prediction assists traffic managers or intelligent systems in making reasonable policies to relieve the impending congestion or giving optimal routes to the travelers. In the practical Sydney Co-ordinated Adaptive Traffic System, the adjustment of traffic signal is supported by the real-time prediction of future traffic condition. In the parallel-transportation management systems, accurate traffic prediction is an essential precondition for building the simulation system that is running in parallel with the actual transportation systems [1]. To settle this issue, a great number of prediction models have been proposed based on historical data or spatio-temporal correlation [2, 3].

In recent years, some spatio-temporal correlation methods have been proposed for urban traffic flow prediction [3–6]. Most of the prevalent approaches pay attention to building the relationship between the adjacent upstream and downstream traffic states. However, in practice, the change of traffic state at the target road is induced not only by its adjacent roads but also by the roads that are not connected

*Correspondence to: Hui Chen, School of Information Science and Engineering, Shandong University, Jinan 250100, China. E-mail: huichen@sdu.edu.cn

with it directly. On the other hand, from the perspective of statistics, abundant data can provide more useful information for the prediction of the response. Therefore, in this paper, we propose to predict the short-term traffic volume on the target road accurately by considering the traffic states from all available road segments in the road network. Towards the high-density or large road network, the number of variables fed into the prediction model can be high-dimensional. Consequently, we propose a spatio-temporal variable selection-based support vector regression (VS-SVR) model fed with the high-dimensional traffic data.

The proposed traffic prediction framework can be presented as a two-stage model. In the first stage, the multivariate adaptive regression splines (MARS) model is employed to select a set of predictors most related to the target from the high-dimensional spatio-temporal variables, and different weights are assigned to the selected predictors. Afterwards, SVR is trained on the weighted variables in the second stage for prediction. In the experiments, the actual traffic volumes collected from a sub-area of Shanghai, China, every 10 minutes are employed to evaluate the proposed model. Some common or state-of-the-art traffic flow prediction methods are also implemented for comparison in our study, including the auto-regression (AR), MARS [7], SVR, seasonal autoregressive integrated moving average (SARIMA), and spatio-temporal Bayesian MARS (ST-BMARS) models [6].

The main contributions of this paper could be summarized in the following three aspects: (i) unlike the mainstream spatio-temporal prediction models, the spatio-temporal correlation in the proposed VS-SVR model is not limited to the adjacent upstream and downstream road segments. VS-SVR is built on the traffic data from the surrounding road network; (ii) a spatio-temporal variable selection-based framework is proposed to reduce the dimension of the independent variables and generate accurate prediction results; (iii) in the experiments, we compare the proposed model with some state-of-the-art models on actual traffic data. The experimental results indicate that the proposed model performs much better than the referenced approaches in view of root mean square error (RMSE) and mean absolute percentage error (MAPE) on most road segments.

The rest of this paper is organized as follows. Section 2 reviews the related works briefly. Section 3 describes the details of the proposed spatio-temporal VS-SVR model. Then, we introduce the actual traffic data used in our work in Section 4. The spatio-temporal correlation and the short-term traffic volume prediction results are illustrated and analyzed in Section 5. Moreover, the sensitivity of VS-SVR is also discussed in this section. Finally, some concluding remarks and directions for the future works are given in Section 6.

2. RELATED WORKS

Historically, great progress has been made in the traffic flow prediction. In general, the prevalent methods can be classified into three categories: univariate time series methods, multivariate spatio-temporal-based methods, and the macroscopic traffic model-based methods.

The univariate time series-based traffic prediction methods build the relationship between the historical and future traffic states at one location or road with parametric or non-parametric models. Among those methods, the autoregressive integrated moving average (ARIMA) is one of the most frequently used parametric methods [8–10]. Even in recent years, the SARIMA is still a state-of-the-art prediction method in many experiments [11]. Meanwhile, the non-parametric models also play important roles in traffic data modeling and prediction as they are flexible to the nonlinear process. Methods that have successfully applied non-parametric regression include k-nearest neighbor [12–14], artificial neural network [15, 16], regression trees [17], the stochastic differential equation [18], the Gaussian processes [19], and deep belief networks [20].

Afterwards, some aggregation approaches were proposed to improve the traffic flow prediction [21]. They combined the linear and non-linear models together to capture the two properties of traffic flow, that is, long-term stability and the short-term instability. Although the time series-based approaches have limited information, some researchers still work on such methods because of its low-cost computation. For example, Tchrakian *et al.* [22] developed a spectral analysis approach for real-time traffic flow prediction. In summary, owing to only considering the temporal correlation of traffic flow, most of these approaches could perform well in the freeway or sparse road network, instead of the dense road network in an urban area.

For the urban road network, the changes of traffic states are impacted by various factors in the time and spatial domain. Consequently, researchers promoted multivariate traffic prediction models by

DOI: 10.1002/atr

introducing the spatial traffic characteristics. For instance, Hobeika and Kim [23] addressed this issue with the aid of traffic states from upstream roads and the average states. Yin *et al.* [24] predicted the traffic flow in current road with the adjacent upstream traffic data using a fuzzy-neural model. Pan *et al.* [5] developed a stochastic cell transmission model based on spatio-temporal correlation for traffic state modeling and prediction. Besides, some other advanced statistical learning approaches have also been extensively exploited to build the relationship between the upstream and downstream traffic states, such as the SVR [25–27], Bayesian graphical model [28], and recurrent neural network [29].

In recent years, a macroscopic urban traffic network (UTN) model was developed to predict the short-term traffic flow in urban area [1, 30]. In contrast with data-driven models, UTN model was built based on the road network model and vehicle moving mechanism. Such model predicts the future traffic state with the current traffic states in the road network other than the historical information. However, the UTN model can only work on the simulation traffic data at present. Therefore, spatio-temporal correlation methods are still generally adopted.

The univariate and multivariate models are data-driven methods, which means such models are trained based on the historical data. The prediction ability of data-driven model is influenced by the quality of training data, the architecture of the model, and the parameters estimation methods. The univariate time series methods are only based on the historical information at the target road. The multivariate spatio-temporal-based methods utilize more information from surrounding roads, while the macroscopic traffic model-based methods are built on the traffic model and the structure of the road network. Such model infers the future traffic state using the current states in the entire road network, including the traffic volume, vehicle queue length, signal setting, and turning rate.

This paper proposes a novel variable selection-based prediction framework, which is a kind of multivariate spatio-temporal-based method. We consider the spatial information from surrounding road segments of the target to build the spatio-temporal correlation. Some recent studies provide comprehensive approaches for incorporating spatial traffic data in prediction models. Lippi *et al.* [31] have developed a "collective" approach to forecasting the traffic flow at multiple nodes of simultaneously using statistical relational learning method. In their work, they have taken into account the relational structure of the spatio-temporal data. However, they tested the model on highways. The urban area we studied is more challengeable. Furthermore, spatio-temporal ARIMA models were also proposed to improve the prediction of downstream locations by feeding on data from upstream detectors [32, 33]. But the spatial structure of the model is specified manually, while in this paper, we learn the spatial correlation with historical data. Vlahogianni *et al.* [4] proposed a modular neural network predictor that integrated traffic volumes from different locations and provided promising results. But these models did not consider the downstream locations of the target. In urban area, we believe that the downstream locations also produce important impact on the upstream traffic situations. In particular, Lippi *et al.* [31] and Vlahogianni *et al.* [4] predicted the traffic states at multiple locations collectively, while in our work, we predict the target roads in the network individually.

The high-dimensional traffic data are selected using MARS model. MARS has been adopted in traffic flow prediction owing to its adaptability to non-linear process. Ye *et al.* [34] predicted the urban traffic volume using MARS based on the spatio-temporal information at the intersections. Xu *et al.* [7] applied MARS model to freeway traffic volume prediction. Afterwards, Xu *et al.* [6] developed a spatio-temporal Bayesian MARS (ST-BMARS) model to estimate the parameters of the prediction model via Bayesian inference and predicted traffic flow competitively in freeway, while in this paper, MARS is only used as a reliable spatio-temporal variable selection approach.

3. METHODOLOGY

Variable selection plays an important role in many applications for which data sets with tens or hundreds of variables are available in classification and regression problems. According to Guyon and Elisseeff's survey [35], the major contributions of variable selection are summarized as follows: improving the prediction performance of the predictor, saving computing resources of the prediction methods, and providing a better understanding of the relationship between the predictors and the response.

Towards the urban traffic prediction, the spatio-temporal correlation of the road segments in a road network is more complicated. Although the road segments are not directly adjacent, the traffic states on

DOI: 10.1002/atr

these segments may interact. Hence, in our research, the current and historical traffic states from all available road segments are collected as the input variables of the prediction model. Consequently, the independent variables can be high-dimensional, especially when the road network is large or high density. The prediction of the traffic flow at a target road segment can be treated as a regression problem with high-dimensional input. Therefore, we propose to obtain better prediction through applying variable selection on such high-dimensional regression problem.

In this paper, we implement the variable selection using a nonlinear and adaptive model, MARS. MARS is an embedded variable selection method as it performs variable selection via a built-in mechanism [36]. The prediction ability of MARS is uncompetitive in contrast with some advanced statistical learning models. Therefore, in our work, MARS is used as a filter that selects useful variables before a learning-based prediction method, SVR. The framework of our proposed urban traffic prediction method is illustrated in Figure 1. Firstly, the predictors $\boldsymbol{X}$ and response $\boldsymbol{y}$ are normalized into $[-1, 1]$ with zero mean and unit variance. The center and scale values in the normalization are preserved and will be used in the prediction of the testing data set. The details of the spatio-temporal variable selection and prediction phases are described in the following content.

3.1. Spatio-temporal variable selection

As we feed the traffic states from all of the road segments at different time intervals into the prediction model, the independent variable $\boldsymbol{X}$ fed into the prediction model is high dimensional. Suppose that we have N observations, $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=i}^{N}$, where $\boldsymbol{x}_i \in \Re^p$ contains p predictors; the corresponding response variables are $\boldsymbol{y} = \{y_i\}_{i=1}^{N}$. To select a subset $\boldsymbol{X}'$ from $\boldsymbol{X}$ that can obtain best prediction performance, assume that the response is generated by

$$y_i = g(\boldsymbol{x}_i) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N\left(0, \delta^2\right) \tag{1}$$

where $\boldsymbol{\epsilon}_i$ is the normal error with a mean of zero and a variance of δ^2. If we define $e(\boldsymbol{X})$ as the estimation error function expressing the error between $\boldsymbol{y}$ and $g(\boldsymbol{X})$, the best predictor set $\boldsymbol{X}'$ should obey the following condition:

$$e\left(\boldsymbol{X}'\right) \leq e\left(\boldsymbol{X}''\right) \tag{2}$$

where $\boldsymbol{X}''$ is any other possible subset of $\boldsymbol{X}$ except $\boldsymbol{X}'$. Hence, the critical issues in the variable selection are twofold: the fitting method and the definition of the error function. Considering that MARS is able to meet the two aspects, we employ MARS to select significant predictors from the spatio-temporal independent variables.

3.1.1. A brief introduction to MARS

Multivariate adaptive regression splines is a flexible multivariate regression model proposed by Friedman [36]. The objective function $g(\boldsymbol{x})$ in MARS is assumed to be composed of a series of basis functions, each of which has its support on a distinct region [39]. The objective function is as shown in the following equation:

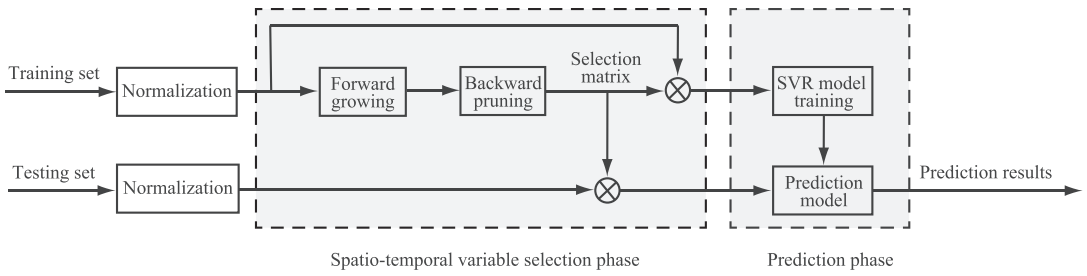

Figure 1. The framework of the proposed urban traffic prediction model.

DOI: 10.1002/atr

$$g(\boldsymbol{x}) = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(\boldsymbol{x}) \tag{3}$$

where β_0 is a constant bias, β_m is the regression coefficient estimated to yield best fit to the desired relationship between the predictor and response, $B_m(\boldsymbol{x})$ is the basis function, and M is the number of basis functions. Generally, $B_m(\boldsymbol{x})$ is expressed as the product of spline functions:

$$B_m(\boldsymbol{x}) = \prod_{l=1}^{L_m} \varphi_{m,l}\left(x_{v(m,l)}\right) \tag{4}$$

where L_m is the degree of interaction of basis B_m and $v(m, l)$ is the index of the predictor variable depending upon the mth basis function and the lth spline function. For each m, $B_m(\boldsymbol{x})$ can consist of a single or a product of two or more spline functions, and no input variable can appear more than once in the product. These spline functions usually take the form as follows:

$$\varphi_{m,l}\left(x_{v(m,l)}\right) \in \left\{ b_q^+\left(x_{v(m,l)} - t_{m,l}\right), b_q^-\left(x_{v(m,l)} - t_{m,l}\right) \right\} \tag{5}$$

with

$$b_q^+(x-t) = [+(x-t)]_+^q = \begin{cases} (x-t)^q, & \text{if } x > t \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$b_q^-(x-t) = [-(x-t)]_+^q = \begin{cases} (t-x)^q, & \text{if } x < t \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $[.]_+$ denotes the positive part of the argument, x is the predictor split, t is the threshold on the predictor, named knot, and q is the power of the spline function.

3.1.2. Variable selection in MARS model

The "optimal" variable selection from $\boldsymbol{X}$ in MARS is reached through a two-stage process: forward growing and backward pruning. In the forward growing stage, MARS initially constructs an over-fitted model by adding an excessively large number of basis functions. These basis functions are allowed to interact with one another or be restricted to entry as additive components only. During this stage, we use a greedy algorithm to consider whether to add a basis function to the candidate model. The basis functions are selected from the initial set randomly and uniformly. Meanwhile, given a configuration for $B_m(\boldsymbol{x}_i)$, the coefficients β_m are estimated by minimizing the residual sum of squares (RSS) criterion with the following form:

$$RSS(M) = \sum_{i=1}^{N} \left(y_i - \hat{g}_M(\boldsymbol{x}_i)\right)^2 \tag{8}$$

where $\hat{g}_M(\boldsymbol{x}_i)$ is the estimation of y_i with M basis functions. Finally, new pairs of functions are considered at each phase until the change of residual error is too small to continue or until the maximum number of basis functions specified at the beginning is reached [36].

In the second backward pruning stage, basis functions are deleted according to their contributions to the model, that is, the terms with negative contribution are deleted, until an "optimal" model is reached. The backward removal is performed by suppressing the basis functions that contribute to a minimal residual error. This stage attempts to increase the generalizability of the model via reducing its complexity. This process can be conducted by means of generalized cross-validation (GCV):

DOI: 10.1002/atr

$$GCV(M) = \frac{\sum_{i=1}^{N}(y_i - g_M(\boldsymbol{x}_i))^2}{(1 - C(M)/N)^2} \tag{9}$$

where M is the number of linearly independent basis functions in Equation (3), being proportional to the number of basis function parameters, and $C(M)$ is a complexity penalty function to avoid over-fitting and indicates the effective number of parameters in the model. Usually, it is defined as $C(M) = \mu \cdot M$, where μ is the penalizing parameter for adding a basis function. In literature [36] and the R package *earth* [37], μ is suggested to be a small integer. The appropriate value of μ will be determined by experiment. Finally, the basic choice of MARS has $\hat{M} = \arg\ \min_M GCV(M)$ additive terms.

After building an "optimal" MARS model through the aforementioned two stages, the importance of the variables can be estimated via checking the variation of GCV when the individual predictor is removed from the model. If a predictor (including spatial and temporal traffic volume) was rarely or never used in any MARS basis function, it has little or negative influence on the target road segment. In our study, the GCV changes are normalized in the [0, 1] interval, where 1 denotes that the predictor is the most important one among all of the predictors, while 0 denotes that the predictor is useless or has negative effect to the response. Finally, the GCV changes of all predictors constitute the variable selection weight $\boldsymbol{\omega}$.

3.2. Support vector regression model

In the second stage of our prediction framework, SVR is employed to learn the relationship between selected spatio-temporal predictors and response. In recent years, SVR has shown remarkable generalization abilities in the prediction of traffic flow [25, 26].

Given the training data set $\{(\boldsymbol{x}_i, , y_i)\}_{i=1}^{N}$, q variables are selected from $\boldsymbol{X}$ by

$$\boldsymbol{x}_i^{'} = \boldsymbol{\omega} \cdot \boldsymbol{x}_i \tag{10}$$

where $\boldsymbol{\omega}$ is the variable selection matrix generated using MARS and $\boldsymbol{x}_i^{'} \in \Re^q$. Then a kernel function $\varphi(\boldsymbol{x}_i^{'}) = \{\varphi_1(\boldsymbol{x}_i^{'}), \ldots, \varphi_D(\boldsymbol{x}_i^{'})\}^T$ is employed to transform the q-dimensional input into a higher D-dimension Hilbert space. Based on the nonlinear mapping, SVR model can be expressed as follows:

$$f\left(\boldsymbol{x}_i^{'}\right) = \boldsymbol{w}^T \varphi\left(\boldsymbol{x}_i^{'}\right) + b \tag{11}$$

where $\boldsymbol{w} = \{w_1, w_2, \ldots, w_D\}^T$ is a weight vector and b is bias.

In SVR, Vapnik [38] introduced ϵ-insensitivity loss function to ignore errors of size less than ϵ. The loss function is defined as follows:

$$L_\epsilon\left(f\left(\boldsymbol{x}_i'\right) - y_i\right) = \max\left\{0, \left|f\left(\boldsymbol{x}_i'\right) - y_i\right| - \epsilon\right\} \tag{12}$$

When the predicted value falls outside the band area, the loss is equal to the difference between the predicted value and the margin. Otherwise, the function gives zero loss. Such error function is therefore more tolerant to noise and is thus more robust. As in the hinge loss, there is a region of no error, which causes sparseness.

After introducing the loss function, $\boldsymbol{w}$ and b are obtained by minimizing the following equation:

$$E(\boldsymbol{w}) = \frac{1}{2}\boldsymbol{w} \cdot \boldsymbol{w} + C\frac{1}{N}\sum_{i=1}^{N} L_\epsilon\left(f\left(\boldsymbol{x}_i'\right) - y_i\right) \tag{13}$$

where C is a parameter controlling the trade-off between the penalty and margin. The issue of SVR is to find the acceptable function $f(\cdot)$ corresponding to a minimum $E(\boldsymbol{w})$. Analogous to the soft margin hyperplane, Vapnik [38] introduced slack variables to account for deviations out of the ϵ-zone. Thus,

DOI: 10.1002/atr

Equation (13) can be transformed to

$$\begin{cases} \min & \frac{1}{2}(\boldsymbol{w}\cdot\boldsymbol{w}) + C\frac{1}{N}\sum_{i=1}^{N}\left(\xi_i + \xi_i^{'}\right) \\ \text{s.t.} & y_i - \left[\boldsymbol{w}\cdot\varphi\boldsymbol{x}_i^{'} + b\right] \le \epsilon + \xi_i^{'} \\ & \left[\boldsymbol{w}\cdot\varphi\boldsymbol{x}_i^{'} + b\right] - y_i \le \epsilon + \xi_i \\ & \xi_i, \xi_i^{'} \ge 0, \quad i = 1, 2, \ldots, N \end{cases} \tag{14}$$

For convenient calculation, Equation (14) is often solved through its dual problem:

$$\begin{cases} \max & -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\beta_i - \beta_i^{'}\right)\left(\beta_j - \beta_j^{'}\right)K\left(\boldsymbol{x}_i^{'}, \boldsymbol{x}_j^{'}\right) \\ & \quad - \epsilon\sum_{i=1}^{N}\left(\beta_i + \beta_i^{'}\right) - \sum_{i=1}^{N}\left(\beta_i - \beta_i^{'}\right) \\ \text{s.t.} & 0 \le \beta_i \le C/N, \; 0 \le \beta_i^{'} \le C/N, \\ & \sum_{i=1}^{N}\left(\beta_i - \beta_i^{'}\right) = 0, \quad i = 1, 2, \ldots, N \end{cases} \tag{15}$$

where β_i and $\beta_i^{'}$ are the Lagrange multipliers of two constraints in Equation (14) and $K\left(\boldsymbol{x}_i^{'}, \boldsymbol{x}_j^{'}\right) = \varphi\left(\boldsymbol{x}_i^{'}\right)\cdot\varphi\left(\boldsymbol{x}_j^{'}\right)$ is the kernel function. After the aforementioned transformations, the solution to Equation (13) is transformed into the estimation of β_i, $\beta_i^{'}$, and b. Finally, in the testing set, we can predict the response of $\boldsymbol{x}'$ by

$$f\left(\boldsymbol{x}^{'}\right) = \sum_{i=1}^{N}\left(\beta_i - \beta_i^{'}\right)K\left(\boldsymbol{x}^{'}, \boldsymbol{x}_i^{'}\right) + b \tag{16}$$

4. DATA SET DESCRIPTION

The work in this paper focuses on the short-term prediction of traffic volume in Shanghai urban road network. The raw data set was collected from Sydney Co-ordinated Adaptive Traffic System and provided by Shanghai Traffic Information Center together with the precise traffic map, as the traffic states in short intervals (2 and 5 min) in urban road network are prone to be interrupted by traffic signal light. The selected field data are the total numbers of vehicles passing certain loop detectors during every interval of 10 minutes, whose unit is vehicles per hour (veh/h). That is, there are 144 data sample points on each day. The duration of the data is from February 1–29, 2012, except February 24, because there were no data on this day. Because the urban traffic states on weekdays are prone to be congested and more complicated, we only consider the traffic volume on weekdays in the experiment.

The traffic volumes in a sub-area of Shanghai road network are investigated in our study. The map and the location of the sub-area are shown in Figure 2. The sub-area consists of 12 bidirectional road segments labeled by the capital letters. The average length of the segments is about 400 m. Furthermore, the footnotes 1 and 2 are used to denote the direction of the vehicle stream. Footnote 1 denotes the traffic streams from north to south or west to east, and footnote 2 denotes the reverses. For example, D_1 denotes the link between its upstream roads, A_1, C_1, F_2, and downstream roads, B_2, E_1, G_1, and flows from west to east; G_2 denotes the link between its upstream roads, I_1, L_2, J_2, and downstream roads, B_2, D_2, E_1, and flows from south to north.

DOI: 10.1002/atr

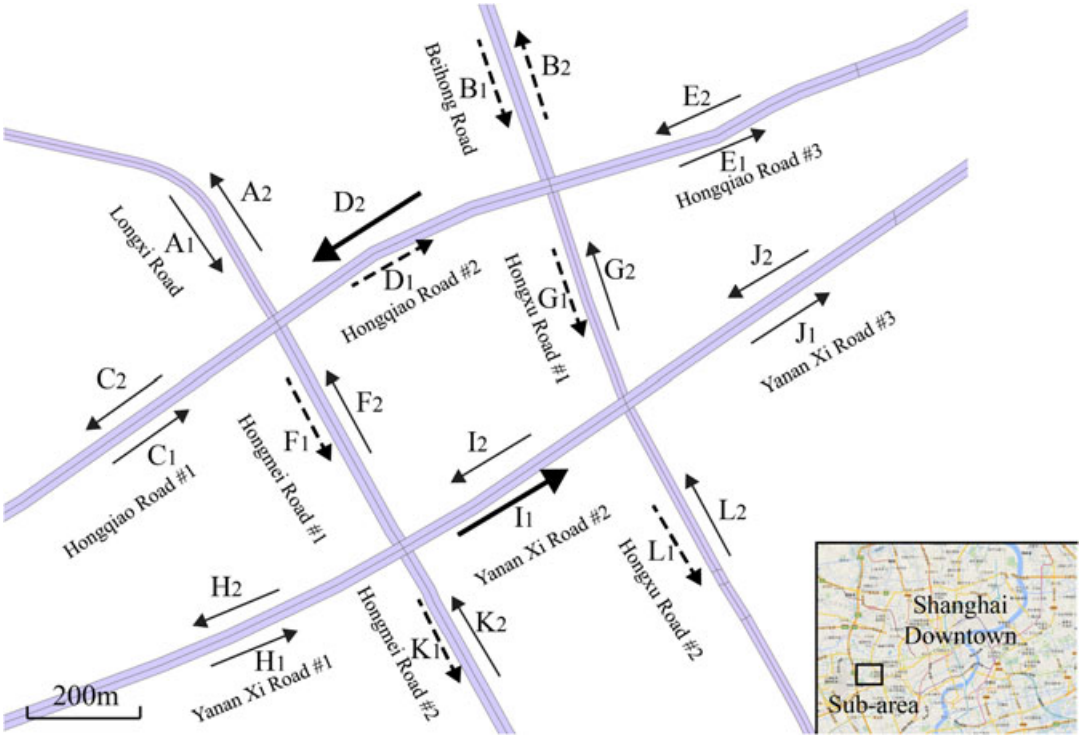

Figure 2. A sub-area road network of Shanghai and the location of the sub-area.

As shown in the map, there are 24 road segments in the selected road network (we consider the bi-directional roads as two individual road segments). However, in practice, some loop detectors do not work or there are no detectors on some road segments. In our data set, traffic data from road B_1, B_2, D_1, F_1, G_1, K_1, and L_1 were unavailable. Therefore, we feed the traffic volume from the 17 available road segments into the proposed prediction model.

Obviously, missing data impacts the performance of prediction models distinctly. Generally, the missing data in our raw data set can be classified into two categories: (i) section missing data, which means data are missing in a long term, for example, several days. The section missing data exists in the seven of the aforementioned road segments. In order to minimize the impact of such missing data to the comparison of models, these road segments without data or with broken loop detectors are excluded in the experiment data; (ii) accidental missing data, which means data is missing at a specific time interval or a few continuous time intervals. In the experiments, the accidental missing data are eliminated when selecting the significant predictors and training the prediction model. In the model testing stage, such missing data are replaced using their predicted values.

5. EXPERIMENTS AND DISCUSSIONS

To build the proposed prediction model, the available data set is divided into two subsets: training set and testing set. The training set contains the early 17 weekdays in our data set, from February 1–23, 2012. In the training set, 32 samples were accidentally missed. The remaining three weekdays, February 27, 28, and 29, are used for model evaluation. The training set is used to obtain the contribution weight $\boldsymbol{\omega}$ via MARS, and then the weighted spatio-temporal variables in the training set are reused to train SVR. Afterwards, the testing set is employed to evaluate the performance of our traffic prediction model.

Predictors from each road segment contain the current traffic volume V_t and the traffic volume in earlier d time intervals, $\{V_{t-1}, \ldots, V_{t-d}\}$. To determine the number of earlier time intervals d, Xie *et. al.* employed an autocorrelation function method [19]. However, previous spatio-temporal models on urban traffic prediction usually construct the independent variable $\boldsymbol{X}$ with small time intervals [28]. In our experiments, we tested the effects of several time intervals, for example, 2, 3, 4, 5. After these

DOI: 10.1002/atr

attempts, we found that the proposed model can reach better prediction results with three time intervals. Consequently, four predictors, $\{V_t, V_{t-1}, V_{t-2}, V_{t-3}$, are drawn from each road segments and then compose the independent variable $\boldsymbol{X}$. Therefore, the independent variable $\boldsymbol{X}$ in the training and testing set both contain 68 predictors. The response is the average traffic volume at the target road in the later 10 minutes, V_{t+1}.

Moreover, in SVR model, the kernel function used for the traffic prediction is $K(x_i, x_j) = \exp(-\sigma \| x_i - x_j \|^2)$, where σ is the kernel parameter. In the SVR training process, the parameters $\{C, \sigma, \epsilon\}$ are selected by a fivefold cross-validation on the training set. The searching grid over $\{C, \sigma, \epsilon\}$ is $[2^{-2}, 2^{-1}, \ldots, 2^6] \times [2^{-6}, 2^{-5}, \ldots, 2^2] \times [2^{-5}, 2^{-4}, \ldots, 2^2]$.

In this section, the variable selection results at two typical road segments with different directions, I_1 and D_2, are illustrated first. As shown in Figure 2, I_1 is a three-lane road segment directing from suburb to downtown. D_2 is a two-lane road segment directing from downtown to suburb. Afterwards, we compare the prediction results of VS-SVR model with some existing prediction models. Besides, the model sensitivity is also inspected at road I_1 via changing the key parameters.

5.1. Spatio-temporal variable selection results

In the variable selection phase of VS-SVR, to simplify the greedy growing process, we defined the stopping criterion in the experiments. The forward growing process stops adding new basis functions when any one of the following conditions is met:

(1) Reaching the maximum number of the basis functions $M_{\max}$.
(2) Adding a term changes RSq by less than a small θ_{RSq}, where RSq is a measure of goodness of fit and is defined as follows:

$$RSq = 1 - \frac{RSS}{(N-1) \times var(\boldsymbol{y})} \tag{17}$$

(3) Reaching a large RSq less than 1, such as 0.999.
(4) Reaching an "optimal" model with maximum accuracy (adding any new term cannot increase RSq).

During the second pruning stage in Section 1, the descent of GCV importance indicates the contribution of the spatio-temporal variables to the target road segment. After the two-step optimization of MARS, the contribution weight $\boldsymbol{\omega}$ can be drawn from the descent of GCV importance. Finally, the weights of the selected predictors for I_1 and D_2 are illustrated in Figure 3a and b, respectively. The predictors that are not presented in the figure are regarded to have negative impacts on the target and will not be fed into the following SVR model.

In Figure 3a, two predictors most related to the response $V_{I_1,t+1}$ are its current state $V_{I_1,t}$ and the state from one of its upstream road segments, $V_{K_2,t}$. Meanwhile, the other predictors selected by MARS model are traffic states on road A_1, J_2, F_2, and so on. Among them, A_1 is the non-adjacent upstream road, J_2 is the reverse road of the downstream road J_1, and F_2 is one of the output roads that belong to the same intersection with I_1.

As shown in Figure 3b, for road D_2, the predictors selected via MARS model are not only its current state $V_{D_2,t}$ but also the traffic states from road A_1, H_2, and so on, where A_1 is one of the input roads that belong to the same intersection with D_2 and H_2 locates at the downstream of D_2. It is worth noting that the impact from upstream roads E_2 and G_2 is less than A_1 and H_2. Therefore, the figures indicate that some road segments would influence on the future traffic states of the target road, although they do not belong to the adjacent upstream or downstream roads.

5.2. Prediction results analyses

Following the spatio-temporal variable selection, the SVR model is built based on the selected predictors with different weights. To verify the prediction ability of the proposed VS-SVR model, the following five traffic prediction models were implemented and evaluated in the experiments:

DOI: 10.1002/atr

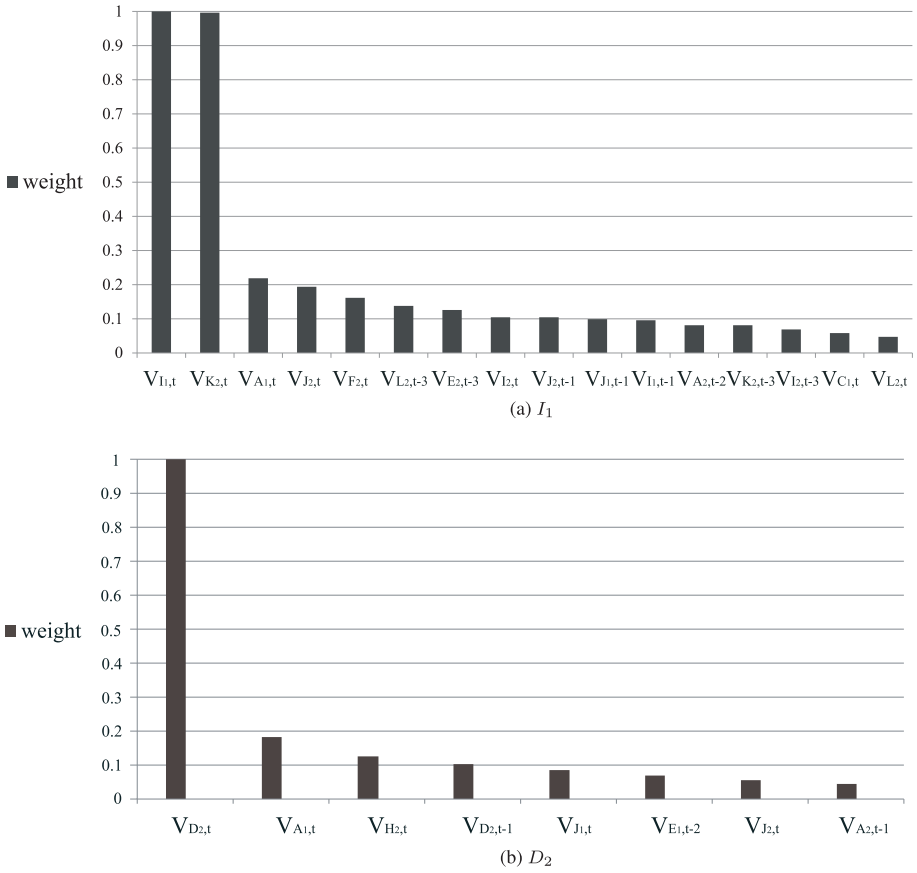

Figure 3. The selected spatio-temporal variables for roads $\boldsymbol{I}_1$ and $\boldsymbol{D}_2$.

(1) AR: AR is widely applied in traffic flow prediction as a time series modeling tool. In most studies, AR served as a simple baseline to appraise the prediction ability of the new models [6, 28]. In this paper, the three-order AR was built to predict the testing set using the temporal traffic states.
(2) MARS: Despite that MARS is used as a variable selection technique in this paper, it can be employed to predict the short-term traffic flow [7]. To demonstrate the promotion of the proposed combined model, MARS was carried out on the testing data set.
(3) SVR: As the prediction module of VS-SVR, SVR was also implemented directly on the traffic states without variable selection. The comparison between SV-SVR and SVR will intuitively reflect the contribution of variable selection in the combined model.
(4) SARIMA: According to the Lippi *et al.* experiments on actual traffic data, SARIMA is still a competitive short-term traffic flow prediction technique at present [11]. Therefore, in our experiments, SARIMA was employed to compare with the VS-SVR model as a state-of-the-art prediction model.
(5) ST-BMARS: ST-BMARS is another state-of-the-art traffic prediction model that gives promising prediction results [6]. Different from SARIMA, ST-BMARS is a spatio-temporal correlation model. Moreover, ST-BMARS works on the data from the road network as VS-SVR does. Consequently, ST-BMARS was considered as a comparable traffic prediction model in our experiments.

Among these referenced models, AR and SARIMA were built on the temporal information; MARS, SVR, and ST-BMARS were built on the same spatio-temporal information with VS-SVR.

As evaluating indicators, two measures for forecasting error analysis, RMSE and MAPE, were employed to evaluate the performance of prediction models. These metrics are commonly used within the field for evaluating model performance. RMSE and MAPE are defined as follows:

DOI: 10.1002/atr

$$RMSE = \sqrt{\left[\frac{1}{K}\sum_{k=1}^{K}\left(V_k - \hat{V}_k\right)^2\right]} \tag{18}$$

$$MAPE = \frac{1}{K}\sum_{k=1}^{K}\frac{\left|V_k - \hat{V_k}\right|}{V_k} \times 100\% \tag{19}$$

where V_k denotes the actual traffic volume and $\hat{V_k}$ is the predicted value produced by the proposed or comparison models. In addition, to minimize the impact of the missing data to the comparison of prediction abilities, the missing data were excluded in the error calculation. Therefore, in Equations (18) and (19), K is the total number of time intervals without the missing data.

In the model testing phase, the relevant six models were carried out on the available 17 road segments during the three weekdays, from February 27 to 29. The prediction results and residuals at each sample point at road I_1 and D_2 during the testing days are plotted in Figures 4 and 5, respectively. As can be observed in these two figures, the traffic volume from road I_1 vibrates much frequently than road D_2. This is due to the fact that the three-lane road usually contains more complex traffic states than two-lane road in the urban area.

In another aspect, the daily trends of traffic flow at roads I_1 and D_2 perform differently in the morning and evening peak. For road I_1, the morning peak is much stronger than the evening peak. In contrast, the evening peak is stronger than the morning peak at road D_2. Such differences are due to the daily origin-destination pattern existing between the suburb and downtown. Generally, most people live in suburb and work in downtown. Therefore, the traffic is comparatively heavy from suburb to downtown in the morning and from downtown to suburb in the evening on weekdays. From the two figures, even though there are different origin-destination trends between I_1 and D_2, the proposed spatio-temporal VS-SVR provides reliable prediction results, especially during the daytime.

Moreover, the prediction results of SARIMA, ST-BMARS, and VS-SVR during the morning peak on February 27 for road I_1 are plotted in Figure 6. As can be seen, the time series models, SARIMA, generated obvious delays when the traffic states change suddenly, for example, the climbing phase at around 7:00 AM (see region R1 in Figure 6). By contrast, the spatio-temporal correlation approaches followed such sudden changes of actual volume timely. Furthermore, observing the region R2 in the figure, there is a precipitate dropping down of the traffic volume at around 9:20 AM, and accurate prediction would be challengeable. Against this situation, only VS-SVR gave satisfying predictions, while SARIMA and ST-BMARS failed. These facts illustrate that the historical information cannot offer sufficient support for the short-term prediction of the traffic state in urban area. The spatial information from surrounding road segments is significant to promote the performance of the prediction models.

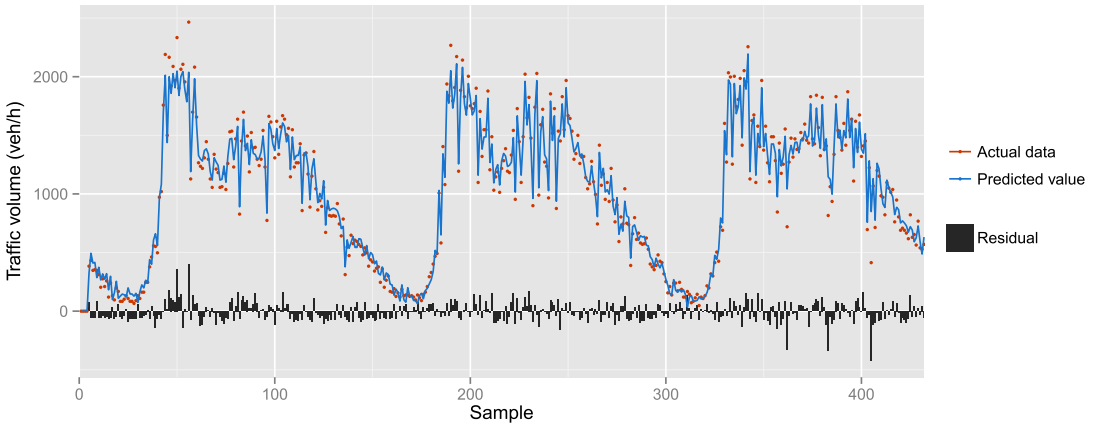

Figure 4. Prediction results of road I_1 on the three testing days.

DOI: 10.1002/atr

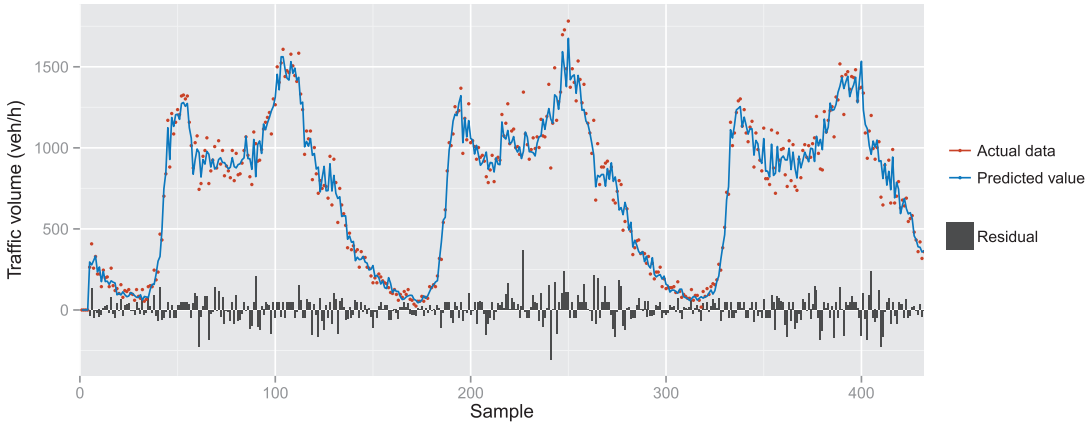

Figure 5. Prediction results of road $\boldsymbol{D}_2$ on the three testing days.

Furthermore, the RMSE and MAPE of the proposed and comparison prediction models are summarized in Tables I and II for road I_1 and D_2, respectively. According to the U.S. Federal Highway Administration (FHWA) quality standards, the maximum acceptable prediction error is 20 %; 10 % should be an ideal error. From this perspective, AR shows an unsatisfactory predictions on our actual urban traffic volume and is slightly better than the maximum acceptable error. SARIMA performs much worse than AR in our experiments because SARIMA mainly relies on the periodic and long-term pattern of the traffic flow. However, the traffic states in urban area are susceptible to the traffic conditions from the surrounding road segments in a short time. In addition, MARS, SVR, and ST-BMARS perform a little better than AR at road I_1. These models do not utilize the spatial information adequately in the view of MAPE. For road D_2, even though these three models perform much better, the proposed VS-SVR is still the best one among the six models in view of RMSE and MAPE.

In addition, the prediction errors of VS-SVR at the remaining 15 road segments are also given in Table III. According to the prediction results of road I_1 and D_2, the time series models perform much worse than the spatio-temporal models. Therefore, in Table III, only the spatio-temporal models are compared. Although at some roads, for example, A_2, C_1, E_1, the promotions are not conspicuous. It can be inferred that VS-SVR could provide better prediction results than MARS and SVR, which are two modules of VS-SVR.

Finally, the performances of VS-SVR and the reference models are summarized as follows:

(1) Rational exploitation of the spatial information could improve the prediction of the traffic flow. This conclusion has been proved by many experiments in current studies. For this reason, the MARS, SVR, ST-BMARS, and VS-SVR models perform much better than AR and SARIMA.

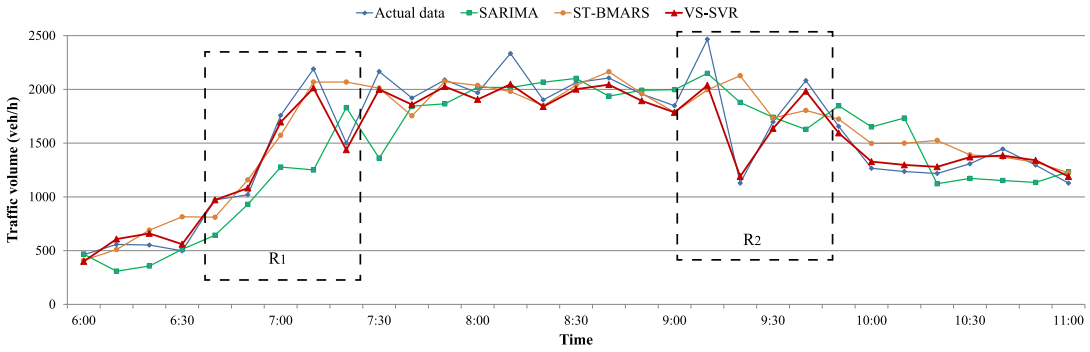

Figure 6. Prediction results comparison of the morning peak at road $\boldsymbol{I}_1$ on February 27. SARIMA, seasonal autoregressive integrated moving average; ST-BMARS, spatio-temporal Bayesian multivariate adaptive regression splines; VS-SVR, variable selection-based support vector regression.

DOI: 10.1002/atr

Table I. Prediction error comparison of the six models at road I_1.

	RMSE				MAPE (%)			
Methods	Feb. 27	Feb. 28	Feb. 29	Total	Feb. 27	Feb. 28	Feb. 29	Total
AR	212.55	221.59	242.97	226.19	19.04	18.01	21.91	19.66
MARS	185.03	206.24	205.19	199.18	19.63	16.13	18.60	18.10
SVR	128.18	111.65	86.85	110.03	17.61	20.50	12.62	16.90
SARIMA	244.43	271.25	267.40	261.30	29.04	30.92	29.07	29.68
ST-BMARS	194.74	207.17	221.49	208.21	19.67	17.82	18.83	18.77
VS-SVR	**75.74**	**59.85**	**77.91**	**71.58**	**12.98**	**9.39**	**10.80**	**11.04**

RMSE, root mean square error; MAPE, mean absolute percentage error; AR, auto-regression; MARS, multivariate adaptive regression splines; SVR, support vector regression; SARIMA, seasonal autoregressive integrated moving average; ST-BMARS, spatio-temporal Bayesian MARS; VS-SVR, variable selection-based support vector regression.

(2) When the input variables are high-dimensional, not all the predictors are "important" for understanding the cause–result relationship of the traffic flows. Some of the predictors play positive roles to the prediction of the future traffic state at the target road. Others may play negative roles to the prediction. As one of the components of VS-SVR, MARS model could select the useful predictors. But one critical drawback existed in MARS is the potential overfit problem. When the trained MARS model is applied on the testing data set, the generalization error might be quite large. In contrast, VS-SVR predicts the response using SVR, which reduces the overfit effectively through introducing a ϵ-insensitivity loss function.
(3) In the other aspect, as the prediction module of VS-SVR, SVR could yield encouraging results in a variety of experiments [11, 25–27]. However, SVR worked under low-dimensional input in these literature. For high-dimensional input in our work, the prediction results of SVR are not persuasive enough. Apparently, the variable selection phase extracts useful predictors for SVR.
(4) As a probabilistic version of MARS, ST-BMARS performs better than MARS in the experiments and is comparable with SVR [6]. Moreover, ST-BMARS also could be used to discover the potential spatio-temporal correlationship between roads in the network. Even so, the prediction ability of VS-SVR outperforms ST-BMARS in our experiments.
(5) Compared with the referenced models, VS-SVR not only yielded better prediction but also discovered the germane road segments surrounding the target. For example, for road I_1, 10 of 17 segments were discovered; for road D_2, 7 of 17 segments were discovered.

To evaluate the goodness of fit of the models, the residual standard deviation could be calculated like this:

$$\hat{\gamma} = \sqrt{\frac{1}{K}\sum_{k=1}^{K} r_k^2} \tag{20}$$

Table II. Prediction error comparison of the six models on road D_2.

	RMSE				MAPE (%)			
Methods	Feb. 27	Feb. 28	Feb. 29	Total	Feb. 27	Feb. 28	Feb. 29	Total
AR	121.50	130.04	126.48	126.10	18.48	18.18	20.21	18.96
MARS	95.82	112.36	100.57	103.32	14.27	15.32	16.21	15.28
SVR	108.68	120.55	107.72	112.50	15.75	15.21	16.64	15.86
SARIMA	135.26	138.48	133.96	135.91	21.55	22.05	21.40	21.67
ST-BMARS	95.88	106.88	97.11	100.11	13.29	13.43	14.62	13.78
VS-SVR	**67.95**	**80.62**	**69.85**	**73.07**	**11.49**	**12.65**	**12.86**	**12.34**

RMSE, root mean square error; MAPE, mean absolute percentage error; AR, auto-regression; MARS, multivariate adaptive regression splines; SVR, support vector regression; SARIMA, seasonal autoregressive integrated moving average; ST-BMARS, spatio-temporal Bayesian MARS; VS-SVR, variable selection-based support vector regression.

DOI: 10.1002/atr

Table III. RMSE of the spatio-temporal models on other roads.

Road ID	RMSE				MAPE (%)			
	MARS	SVR	ST-BMARS	VS-SVR	MARS	SVR	ST-BMARS	VS-SVR
A_1	55.83	47.60	42.73	35.48	24.76	18.39	16.47	15.35
A_2	57.20	44.28	51.59	31.30	22.04	15.62	16.99	14.29
C_1	57.57	48.81	59.84	41.94	16.82	14.31	18.37	13.80
C_2	102.49	90.65	86.99	48.78	27.57	21.33	22.79	18.39
E_1	117.84	99.46	109.71	79.94	23.48	17.69	20.92	16.25
E_2	76.69	67.53	68.59	39.88	22.50	16.13	18.55	14.85
F_2	59.33	48.65	54.69	40.68	20.34	15.28	19.77	13.53
G_2	88.12	74.49	79.58	48.66	21.42	17.74	17.26	14.94
H_1	118.30	108.11	97.34	67.16	23.47	18.92	18.79	12.65
H_2	83.50	69.56	63.80	55.14	23.61	18.41	17.54	14.16
I_2	234.67	226.36	208.82	203.27	19.00	16.06	14.79	10.91
J_1	107.59	90.96	73.05	84.30	21.46	18.16	16.28	14.69
J_2	245.11	185.73	161.24	123.52	16.66	12.00	11.29	9.80
K_2	52.82	50.62	50.52	39.40	28.99	22.46	21.35	13.18
L_2	41.17	35.76	40.47	20.34	25.71	23.56	20.28	15.26
Average	**96.54**	**88.88**	**91.60**	**64.97**	**22.42**	**17.58**	**17.93**	**13.85**

RMSE, root mean square error; MAPE, mean absolute percentage error; MARS, multivariate adaptive regression splines; SVR, support vector regression; ST-BMARS, spatio-temporal Bayesian MARS; VS-SVR, variable selection-based support vector regression.

where the residuals, $r_k = V_k - \hat{V}_k$, are the differences between the actual data and the predicted values. The goodness of fit of the models can be summarized by $\hat{\gamma}$ and R^2, where $R^2 = 1 - \hat{\gamma}^2/s^2$ and s^2 denotes the variance of the traffic volume. The $\hat{\gamma}$ and R^2 of the reference models on the two typical road segments I_1 and D_2 are summarized in Table IV. As can be seen in the table, VS-SVR has the lowest $\hat{\gamma}$ and R^2 for both I_1 and D_2. Moreover, the R^2 are close to 1, which implies that VS-SVR is a promising traffic flow prediction model.

Besides comparing the prediction ability, we also discussed the computation demand of the prediction models. The average times of 10 executions for the model training and testing at road I_1 are listed in Table V. The time is measured on a PC with 2.8 GHz Intel CPU 4 GB RAM, and 64-bit operating system. As the table shows, VS-SVR consumed much more time than SVR in the model training stage because of the time-consuming variable selection phase. Actually, in practice, variable selection could be performed off-line because the spatio-temporal relationship is relatively stable. The prediction phase using SVR could meet the real-time computing. Therefore, our system can be used to first select variables off-line, and then to perform online forecasts. Besides, the training stage of VS-SVR took as long time as the state-of-the-art ST-BMARS on the testing road network. However, the prediction stage of VS-SVR took much less time and could meet the real-time online application.

Table IV. The goodness of fit of the six models.

Models	Road I_1		Road D_2	
	γ	R^2	γ	R^2
AR	211.3	0.855	121.2	0.921
MARS	188.1	0.888	100.9	0.947
SVR	112.55	0.957	109.1	0.937
SARIMA	244.8	0.807	131.7	0.910
ST-BMARS	200.2	0.879	100.2	0.948
VS-SVR	93.98	0.982	71.54	0.974

AR, auto-regression; MARS, multivariate adaptive regression splines; SVR, support vector regression; SARIMA, seasonal autoregressive integrated moving average; ST-BMARS, spatio-temporal Bayesian MARS; VS-SVR, variable selection-based support vector regression.

DOI: 10.1002/atr

Table V. Running time comparison of the six models.

Models	Number of predictors	Training time(s)	Prediction time(s)
AR	4	0.13	0.006
MARS	68	522.47	0.03
SVR	68	3.67	0.19
SARIMA	4	1.38	0.013
ST-BMARS	68	542.63	13.67
VS-SVR	68	534.94	0.09

AR, auto-regression; MARS, multivariate adaptive regression splines; SVR, support vector regression; SARIMA, seasonal autoregressive integrated moving average; ST-BMARS, spatio-temporal Bayesian MARS; VS-SVR, variable selection-based support vector regression.

Based on the previous experiment results and discussions, we can draw a conclusion that the proposed VS-SVR model can exploit the spatio-temporal correlations reasonably and achieve reliable prediction results in a short time. Moreover, although we tested the model on a small road network, the model also can be easily extended to the large road networks. If, for instance, the object road network is extremely large (e.g., Shanghai road network with thousands road segments), although the model could run directly, a reasonable solution would be to divide the road network into small sub-areas first (e.g., sub-areas with less than 100 road segments). Afterwards, the proposed VS-SVR can be carried out on each sub-area.

5. 3. Model sensitivity analysis

The robustness of the prediction model is another issue need to be concerned besides accuracy and efficiency. Consequently, the model sensitivity to key parameters is examined on the testing data set in this section. In addition, the collinearity of the selected predictors and its influence to the prediction results are also verified in this section.

In the variable selection process, the choice of GCV penalty μ will affect the selection results and their weights. As a result, the final prediction error will vary with the change of μ. In the experiments, μ was changed from 1 to 8 to check the variety of prediction error at road I_1. The RMSE and MAPE of the testing data set with different μ are plotted in Figure 7. When $\mu \geq 3$, the prediction error tends to be steady. Therefore, the GCV penalty μ is set to be 4, which is also consistent with Friedman's suggestion [36].

In the other aspect, the number of selected predictors in the variable selection process will impact the final prediction accuracy. Consequently, we attempted various combinations of some key parameters to generate different number of selected predictors. These parameters include the maximum number of basis functions $M_{\max}$ and the threshold θ_{RSq} in MARS model growing stage and the degree of the

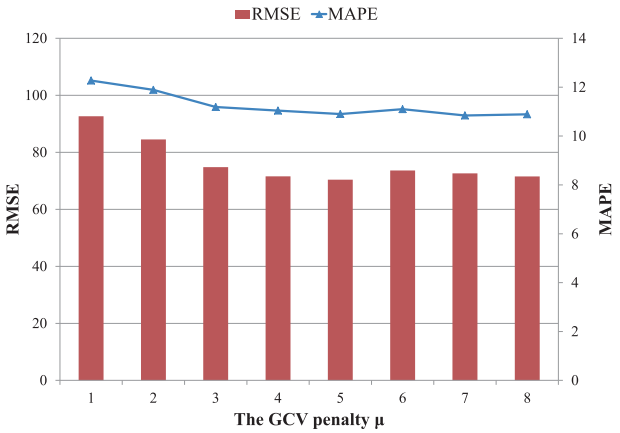

Figure 7. The prediction error with respect to different value of generalized cross-validation (GCV) penalty at road I_1.

DOI: 10.1002/atr

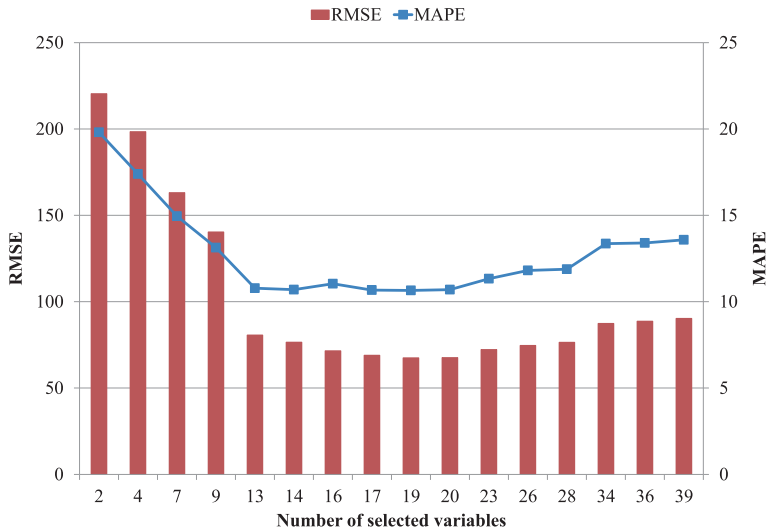

Figure 8. The prediction error with respect to different number of selected variables at road. RMSE, root mean square error; MAPE, mean absolute percentage error.

interaction of basis function L_m in Equation (4). In the experiments, the number of selected predictors N_p ranged from 2 to 39. The prediction errors on the testing data set at road I_1 are plotted in Figure 8. From the figure, it is obvious that

(1) when $N_p < 13$, the RMSE and MAPE drop quickly with the increasing of N_p;
(2) when $13 \leq N_p \leq 20$, the prediction error becomes steady temporarily; and
(3) when $N_p > 20$, the prediction error begins rising again.

Apparently, when N_p increases to 68, which means all predictors are fed into SVR without selection, the VS-SVR model is deduced to SVR model. Its MAPE will increase to 16.90 as presented in Table I.

Furthermore, the model sensitivity to the collinearity between different predictors was also discussed. Different from the factor analysis approaches, MARS may select predictors with collinearity. In our study, the collinearity is measured by the determinant of the covariance matrix of the concerned predictors. The results range from 0 to 1, where 0 indicates that the two predictors are perfect collinear and 1 indicates they are completely independent. Taking road I_1 as an example, the collinearity of the selected 16 predictors are illustrated in Table VI. As can be seen in the table, three predictors, X_3, X_5, X_{15}, have strong collinearity. Consequently, we fed the predictors without the aforementioned

Table VI. Collinearity matrix for selected spatio-temporal variables.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
X_1																
X_2	0.219															
X_3	0.219	0.212														
X_4	0.205	0.413	0.314													
X_5	0.169	0.170	0.116	0.287												
X_6	0.445	0.534	0.475	0.370	0.445											
X_7	0.548	0.618	0.476	0.320	0.491	0.510										
X_8	0.693	0.735	0.679	0.604	0.682	0.762	0.658									
X_9	0.365	0.447	0.337	0.190	0.311	0.381	0.321	0.643								
X_{10}	0.240	0.182	0.201	0.396	0.182	0.518	0.567	0.731	0.413							
X_{11}	0.209	0.237	0.215	0.326	0.178	0.435	0.531	0.715	0.205	0.219						
X_{12}	0.396	0.466	0.307	0.214	0.319	0.413	0.222	0.610	0.198	0.421	0.356					
X_{13}	0.463	0.540	0.418	0.299	0.414	0.437	0.296	0.672	0.267	0.505	0.418	0.248				
X_{14}	0.742	0.784	0.721	0.634	0.725	0.771	0.561	0.327	0.637	0.758	0.736	0.595	0.659			
X_{15}	0.197	0.227	0.150	0.216	0.081	0.403	0.378	0.638	0.230	0.217	0.191	0.226	0.332	0.673		
X_{16}	0.341	0.479	0.437	0.350	0.395	0.354	0.597	0.771	0.406	0.473	0.395	0.457	0.498	0.789	0.376	

DOI: 10.1002/atr

Table VII. Prediction errors without collinear predictors.

	$\boldsymbol{X}'_{(X_3)}$	$\boldsymbol{X}'_{(X_5)}$	$\boldsymbol{X}'_{(X_{15})}$	$\boldsymbol{X}'$
RMSE	72.50	73.18	74.11	71.58
MAPE	11.12	10.80	11.03	11.04

Note: $\boldsymbol{X}'_{(X_p)}$ denotes the independent variables without X_p.
RMSE, root mean square error; MAPE, mean absolute percentage error.

three ones into the SVR model separately. The prediction errors are given in Table VII. The three cases show competitive performance with the original model. The results indicate that the influence of collinearity in the variable selection phase on the final prediction accuracy is minor.

6. CONCLUSIONS

This paper proposed a novel urban traffic volume prediction method based on the spatio-temporal variable selection strategy. The proposed prediction framework contains two critical modules, the variable selection module and the prediction module. In the first stage, MARS is employed as a filter variable selection method. In the second stage, SVR is employed to train the relationship between selected predictors and response. On the other hand, different from most prevalent urban traffic prediction models, the traffic data from all of the road segments are fed into the proposed model other than the adjacent roads of the target. In the experiments, to evaluate the performance of the proposed prediction model, AR, MARS, SVR, and two state-of-the-art models, SARIMA and ST-BMARS, were implemented for comparison. The experiments were carried out on a sub-area of the real road network in Shanghai. The results indicate that the spatio-temporal variable selection-based SVR model is an effective approach for short-term traffic volume prediction in a complex urban road network.

For the future work, the proposed method can be extended in several ways, for example, considering the spatio-temporal correlations under different traffic states, such as morning peak, evening peak, and the stable traffic stage. On the other hand, the balance between the scale of the road network and the accuracy of the model is another significant issue. It would be interesting to find an appropriate size of the road network. Moreover, when predicting the traffic flow in a large-scale traffic network in the future, a reasonable road network partition model should be developed before applying the proposed method.

ACKNOWLEDGEMENTS

This research has been supported by funding from China NSFC under grant 61375019, 61203169, and 61473288; Beijing Natural Science Foundation under grant 4142055; and Natural Science Found of Shandong under grant ZR2013FM032.

REFERENCES

1. Kong Q-J, Xu Y, Lin S, Wen D, Zhu F Liu Y. UTN-model-based traffic flow prediction for parallel-transportation management systems. *IEEE Transactions on Intelligent Transportation Systems* 2013; **14**(3): 1541–1547.
2. Vlahogianni EI, Golias JC Karlaftis MG. Short-term traffic forecasting: overview of objectives and methods. *Transport Reviews* 2004; **24**(5): 533–557.
3. Vlahogianni EI, Karlaftis MG Golias JC. Short-term traffic forecasting: where we are and where we're going. *Transportation Research Part C: Emerging Technologies* 2014; **43**part 1: 3–19.
4. Vlahogianni EI, Karlaftis MG Golias JC. Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks. *Computer-Aided Civil and Infrastructure Engineering* 2007; **22**(5): 317–325.
5. Pan T, Sumalee A, Zhong R Indra-Payoong N. Short-term traffic state prediction based on temporal–spatial correlation. *IEEE Transactions on Intelligent Transportation Systems* 2013; **14**(3): 1242–1254.
6. Xu Y, Kong Q-J, Klette R Liu Y. Accurate and interpretable Bayesian MARS for traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems* 2014; **15**(6): 2457–2469.
7. Xu Y, Kong Q-J, Liu Y. A spatio-temporal multivariate adaptive regression splines approach for short-term freeway traffic volume prediction, in *Proc. 16th International IEEE Conference on Intelligent Transportation Systems*, Oct 2013, pp. 217–222.

DOI: 10.1002/atr

8. Voort MVD, Dougherty M Watson S. Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies* 1996; **4**(5): 307–318.
9. Williams BM, Durvasula PK Brown DE. Urban freeway travel prediction: application of seasonal ARIMA and exponential smoothing models. *Transportation Research Record: Journal of the Transportation Research Board* 1998; **1644**: 132–141.
10. Williams BM. Multivariate vehicular traffic flow prediction: evaluation of ARIMAX modeling. *Transportation Research Record: Journal of the Transportation Research Board* 2001; **1776**: 194–200.
11. Lippi M, Bertini M Frasconi P. Short-term traffic flow forecasting: an experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems* 2013; **14**(2): 871–882.
12. Davis GA, Nihan NL. Nonparametric regression and short-term freeway traffic forecasting. *ASCE Journal of Transportation Engineering* 1991; **17**(2): 168–178.
13. Smith BL, Williams BM Oswald RK. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies* 2002; **10**(4): 303–321.
14. Clark S. Traffic prediction using multivariate nonparametric regression. *Journal of Transportation Engineering* 2003; **129**(2): 161–168.
15. Dougherty MS, Cobbett MR. Short-term inter-urban traffic forecasts using neural networks. *International Journal of Forecasting* 1997; **13**(1): 21–31.
16. Vlahogianni EI, Karlaftis MG Golias JC. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. *Transportation Research Part C: Emerging Technologies* 2005; **13**(3): 211–234.
17. Xu Y, Kong Q-J, LiuY. Short-term traffic volume prediction using classification and regression trees, in *Proc. 2013 IEEE Intelligent Vehicles Symposium*, Gold Coast, Australia, 2013, pp. 493–498.
18. Tahmasbi R, Hashemi S. Modeling and forecasting the urban volume using stochastic differential equations. *IEEE Transactions on Intelligent Transportation Systems* 2014; **15**(1): 250–259.
19. Xie Y, Zhao K, Sun Y Chen D. Gaussian processes for short-term traffic volume forecasting. *Transportation Research Record: Journal of the Transportation Research Board* 2010; **2165**: 69–78.
20. Huang W, Song G, Hong H Xie K. Deep architecture for traffic flow prediction: deep belief networks with multitask learning. *IEEE Transactions on Intelligent Transportation Systems* 2014; **15**(5): 2191–2201.
21. Tan M-C, Wong SC, Xu J-M, Guan Z-R Zhang P. An aggregation approach to short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems* 2009; **10**(1): 60–69.
22. Tchrakian TT, Basu B O'Mahony M. Real-time traffic flow forecasting using spectral analysis. *IEEE Transactions on Intelligent Transportation Systems* 2012; **13**(2): 519–526.
23. Hobeika A, Kim CK. Traffic-flow-prediction systems based on upstream traffic, in *Proc. Vehicle Navigation and Information Systems Conference*, Yokohama, Japan, 1994, pp. 345–350.
24. Yin H, Wong SC, Xu J Wong CK. Urban traffic flow prediction using a fuzzy-neural approach. *Transportation Research Part C: Emerging Technologies* 2002; **10**(2): 85–98.
25. Zhang Y, Xie Y. Forecasting of short-term freeway volume with v-support vector machines. *Transportation Research Record: Journal of the Transportation Research Board* 2007; **2024**: 92–99.
26. Castro-Neto M, Jeong Y-S, Jeong M-K Han LD. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions. *Expert Systems with Applications* 2009; **36**(3Part 2): 6164–6173.
27. Jeong Y-S, Byon Y-J, Mendonca Castro-Neto M Easa S. Supervised weighting-online learning algorithm for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems* 2013; **14**(4): 1700–1707.
28. Sun S, Zhang C Yu G. A Bayesian network approach to traffic flow forecasting. *IEEE Transactions on Intelligent Transportation Systems* 2006; **7**(1): 124–132.
29. Zeng X, Zhang Y. Development of recurrent neural network considering temporal–spatial input dynamics for freeway travel time modeling. *Computer-Aided Civil and Infrastructure Engineering* 2013; **28**(5): 359–371.
30. Xu Y, Kong Q-J, Lin S, Liu Y. Urban traffic flow prediction based on road network model, in *Proc. 9th IEEE International Conference on Networking, Sensing and Control*, Beijing, China, 2012, pp. 334–339.
31. Lippi M, Bertini M, Frasconi P. Collective traffic forecasting, In Machine Learning and Knowledge Discovery in Databases, ser. *Lecture Notes in Computer Science*, 2010, **6322**:259–273.
32. Stathopoulos A, Karlaftis MG. A multivariate state–space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies* 2003; **11**(2): 121–135.
33. Min W, Wynter L. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies* 2011; **19**(4): 606–616.
34. Ye S, He Y, Hu J, Zhang Z. Short-term traffic flow forecasting based on MARS, in *Proc. 5th International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 5, Jinan, China, 2008, 669–675.
35. Guyon I, Elisseeff A. "An introduction to variable and feature selection," *Journal of Machine Learning Research* 2003; **3**:1157–1182.
36. Friedman JH. Multivariate adaptive regression splines. *The Annual of Statistics* 1991; **19**(1): 1–67.
37. S. M. D. from mda:mars by Trevor Hastie and R. T. U. A. M. F. utilities with Thomas Lumley's leaps wrapper., *earth*: *Multivariate Adaptive Regression Spline Models*, 2014, r package version 3.2-7. [Online]. Available: http://CRAN.R-project.org/package=earth. [accessed April 01 2014].
38. Vapnik VN. *The Nature of Statistical Learning Theory*Springer: New York, 1995.
39. Clarke B, Fokoué E Zhang HH. *Principles and Theory for Data Mining and Machine Learning ser. Springer Series in Statistics*Springer-Verlag: Berlin, Germany, 2009.

DOI: 10.1002/atr