Accurate and Interpretable Bayesian MARS for Traffic Flow Prediction

Yanyan Xu, Qing-Jie Kong, *Member, IEEE*, Reinhard Klette, and Yuncai Liu, *Member, IEEE*

***Abstract*—Current research on traffic flow prediction mainly concentrates on generating accurate prediction results based on intelligent or combined algorithms but ignores the interpretability of the prediction model. In practice, however, the interpretability of the model is equally important for traffic managers to realize which road segment in the road network will affect the future traffic state of the target segment in a specific time interval and when such an influence is expected to happen. In this paper, an interpretable and adaptable spatiotemporal Bayesian multivariate adaptive-regression splines (ST-BMARS) model is developed to predict short-term freeway traffic flow accurately. The parameters in the model are estimated in the way of Bayesian inference, and the optimal models are obtained using a Markov chain Monte Carlo (MCMC) simulation. In order to investigate the spatial relationship of the freeway traffic flow, all of the road segments on the freeway are taken into account for the traffic prediction of the target road segment. In our experiments, actual traffic data collected from a series of observation stations along freeway Interstate 205 in Portland, OR, USA, are used to evaluate the performance of the model. Experimental results indicate that the proposed interpretable ST-BMARS model is robust and can generate superior prediction accuracy in contrast with the temporal MARS model, the parametric model autoregressive integrated moving averaging (ARIMA), the state-of-the-art seasonal ARIMA model, and the kernel method support vector regression.**

***Index Terms*—Bayesian inference, interpretable model, Markov chain Monte Carlo (MCMC), multivariate adaptive-regression splines (MARS), spatiotemporal relationship analysis, traffic flow prediction.**

I. Introduction

SHORT-TERM traffic flow prediction is a complex nonlinear but crucial task in intelligent transportation systems (ITS) and has drawn growing attention from many researchers and engineers in the past few decades. It is of basic importance for many components of ITS, such as advanced traffic management systems, adaptive traffic control systems, or traffic information services systems. In the past decade, the short-term traffic flow prediction module has been well exploited in some representative ITS, including the Sydney Coordinated Adaptive Traffic System, the Split Cycle Offset Optimization Technique, and parallel-transportation management systems [1], [2].

From the very beginning of ITS, a great number of scholars and engineers have exploited an extensive variety of mathematical specifications to model traffic characteristics and produce short-term traffic predictions in an equally diverse variety of conditions. Among these traffic prediction methods, apart from some specific methods based on the macroscopic physical model of the road network [2], [3], many methods tried to build parametric or nonparametric data-driven models based on extensive historical traffic data, which have been considered as the most important factor for the prediction model [4].

For instance, researchers have taken advantage of temporal historical data to predict short-term traffic flow through Kalman filtering [5], autoregressive integrated moving averaging (ARIMA) [6], seasonal ARIMA (SARIMA) [7], nonparametric regression methods such as the k-nearest neighbor approach [8], and spectral analysis [9]. These methods can be also regarded as univariate methods as they are fed with the univariate historical values for the modeled road. On the basis of considering the traffic flow as time series, these approaches mostly perform well when the traffic states remain relatively stable, different in more complicated situations.

In recent years, researchers have gradually perceived the significance of spatial information in traffic prediction. Hobeika and Kim [10] tried to predict short-term traffic flow based on current traffic, historical average, and upstream traffic. Sun *et al.* [11] proposed a Bayesian prediction approach taking into account historical data from both current and upstream adjacent segments. Vlahogianni *et al.* [12] exploited a modular neural predictor that was fed with traffic data from sequential locations to improve the accuracy of short-term forecasts. Min and Wynter [13] predicted road traffic by considering the spatial characteristics of a road network, including the distance and the average speed of the upstream segments.

Furthermore, machine learning approaches have also been extensively utilized to deal with short-term traffic flow prediction, such as support vector machines [14], the online support vector regression (SVR) method [15], Gaussian processes [16], and a stochastic approach [17].

Although the previously mentioned spatiotemporal correlation models are quite flexible, they also come with two drawbacks. First of all, most models do not fully exploit the spatial

Manuscript received August 13, 2013; revised January 12, 2014 and March 24, 2014; accepted April 1, 2014. Date of publication May 2, 2014; date of current version December 1, 2014. This work was supported in part by the China National 863 Key Program under Grant 2012AA112307, by the Science and Technology Commission of Shanghai Municipality Program under Grant 11231202801, by the National Natural Science Foundation of China Program under Grant 61104160, and by the Beijing Natural Science Foundation under Grant 4142055. The Associate Editor for this paper was S. Sun.

Y. Xu and Y. Liu are with the Department of Automation and the Ministry of Education of China Key Laboratory of System Control and Information Processing, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xustone@sjtu.edu.cn; whomliu@sjtu.edu.cn).

Q.-J. Kong is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: qingjie.kong@ia.ac.cn).

R. Klette is with the Department of Computer Science, The University of Auckland, Auckland 1020, New Zealand (e-mail: r.klette@auckland.ac.nz).

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

Digital Object Identifier 10.1109/TITS.2014.2315794

1524-9050 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

information collected from the whole road network. Previous spatiotemporal approaches always try to build a specific relationship of the traffic states between the adjacent road segments and the current segment [11]–[13], [18]. The predictors fed into the prediction models are only the traffic states from the adjacent upstream or downstream road segments together with the objective segment. However, other traffic states from road segments or stations, which are not immediately adjacent, are neglected.

Another drawback is that the interpretability of traffic prediction models does not attract sufficient attention in the previous literature. In the practice of traffic control, the interpretability of the prediction model is particularly important. An interpretable model can assist a traffic manager to devise reasonable strategies via extracting the specific road segments that have the maximum contribution on the future traffic state of the target road segment. Although some time-series models, such as the ARIMA model or the regression trees model, are highly interpretable, they can prove to be too rigid when complex nonlinear traffic states are present. In contrast, some more advanced models that were recently proposed appear to be able to export satisfying prediction results, but it has been difficult for the authors to interpret the contribution of each predictor or road segment to the target variable based on available information.

For example, prediction models based on SVR, artificial neural networks, or Gaussian processing are flexible for nonlinear unknown functions, but they lack reasonable interpretation because of their "black box" properties [11], [14], [15]. Therefore, such black box models cannot help a traffic manager to excavate the concrete factors for future traffic states at the target road segments.

Our work aims at addressing the previously mentioned two drawbacks identified in previous work. In this paper, a flexible yet interpretable spatiotemporal Bayesian multivariate adaptive-regression splines (ST-BMARS) model is developed to investigate the relationships between road segments and predict short-term traffic flow accurately. The Bayesian inference implemented through Markov chain Monte Carlo (MCMC) simulation is employed to obtain a series of stable and well-adaptive MARS models in our study.

Moreover, the traffic volume collected from all observation stations on the freeway, including adjacent and nonadjacent stations, are fed into the prediction model and are flexibly selected for volume prediction for the target road segment. In our experiments, actual traffic data collected from a series of observation stations along a freeway in Portland, OR, USA, (every 15 min) are exploited to verify the effectivity of the proposed prediction model. Afterward, relationships between road segments are first investigated by analyzing the importance of variables in the built model. Moreover, three classic and frequently employed methods in previous studies, including the ARIMA, SARIMA, and SVR methods, are briefly reviewed and implemented for comparison with the proposed ST-BMARS model.

The following are the novel contributions in this paper. First of all, the Bayesian MARS model is, for the first time, applied to the traffic prediction problem. Subsequently, a spatiotemporal variant of a Bayesian MARS model is developed for taking full advantage of the traffic data in the road network, including traffic data from upstream and downstream road segments, as well as their historical data. Moreover, we show that the interpretability and the accuracy are well balanced in the proposed prediction model. The interpretability assists traffic managers to find the relationship between the traffic states at a series of observation stations. Meanwhile, its prediction accuracy surpasses some state-of-the-art prediction methods, such as SARIMA and SVR.

The remainder of this paper is structured as follows. Section II states the problem to be solved in this paper and the related work; Section III describes the theory of the ST-BMARS model; Section IV states the traffic data used in our work, the application of the model in practice, and the referenced comparison models; the interpretability, predictive ability, and the robustness of the proposed model are presented and discussed in Section V; and, finally, some concluding remarks and directions for future work are given in Section VI.

II. Problem Statement and Related Work

Short-term traffic flow prediction is a complex nonlinear task, which has been the subject of many research efforts in the past few decades. Researchers have focused on achieving accurate prediction results using various mathematical models. However, in traffic engineering practice, when traffic managers design specific strategies to alleviate the heavy traffic, the interpretability of the traffic prediction model is particularly important. An interpretable prediction model can assist traffic managers to make reasonable strategies by focusing on the most related stations that have the greatest contributions on the future traffic state of the target station. Hence, in this paper, we desire to find these most related stations via an accurate and interpretable prediction model.

Although the interpretability of the prediction models was seldom raised in the literature, some models are interpretable, particularly the parametric models. ARIMA is the most frequently used parametric model and performs well in practice. ARIMA builds the relationship between the past few traffic states and the future state and can provide a clear causality in time domain [8]. The SARIMA model improves the predictive accuracy via drawing the periodicity of the traffic data. It constructs the independent variables using the traffic data in the past several intervals together with the historical data in the same intervals in the last week [7], [19]. SARIMA not only provides the short-term causality but also the long-term change rule of traffic state. Although these models indicate the relationship between the response and the historical data intuitively, they still only work in time domain.

Afterward, some interpretable spatiotemporal models are proposed. Kamarianakis and Prastacos [20] employed the space–time ARIMA (STARIMA) to model the traffic flow in road network and constructed the weighting matrices on the basis of the distances among the observation locations. However, the model is based on the following assumptions: 1) the effect only depends on the distance between the measurement locations; 2) the traffic flow is stable, and no congestion happened; and 3) the traffic states at downstream locations only depend on upstream locations but not vice versa. These

assumptions are clearly too solid for busy freeways or urban road networks. Min and Wynter [13] addressed a multivariate spatiotemporalautoregressive (MSTAR) model for traffic volume and speed prediction. Their model took into account the spatial characteristics of a road network on the basis of the length and average speed of the links. However, they only considered such spatial effects from the neighboring links. In this paper, we desire to develop an accurate traffic prediction model and derive the contributions of any other stations in the road network to the target one using the model.

Moreover, before developing a data-driven prediction model, three issues are frequently considered: selection of the traffic parameter, resolution of the traffic data, and preprocessing of the missing values.

The most commonly used variables in traffic prediction are the three fundamental macroscopic traffic parameters: volume, occupancy, and speed. In most cases, traffic volume is more easily obtained and relatively accurate. Taking the most common traffic information detection equipment, loop detector, for example, loop detector can obtain the number of passing vehicles, the occupancy, and the speed. However, the occupancy and the speed at a location are more susceptible to the driver's behavior (e.g., slow-moving vehicles in low flow conditions). Therefore, in this paper, the traffic volume is considered as the input parameter into the developed model.

The resolution of the traffic parameter is another important issue, particularly in data-driven models, because it affects the quality of information about traffic conditions lying in the data. In general, data must be available in such a form that captures the dynamics of traffic and can be easily predicted. The Highway Capacity Manual from Transportation Research Board indicates the 15-min interval as the best prediction interval as traffic flow exhibits strong fluctuations at shorter intervals [4]. In our work, the traffic volume are aggregated in 15-min interval and expressed as number of vehicles per lane per hour (VPLPH).

Furthermore, in practice, traffic data usually include missing values resulting from the malfunction of the data collection and transmission mechanisms. Before building the prediction model, we eliminate the missing values from the training data set. When we perform the prediction model on the testing data set, the missing data are replaced using their predicted values.

After the traffic data are prepared, the independent variables and the response should be defined. Let $\mathbf{s}_{j,t}$ be the traffic state vector in time interval t at the jth observation station in the road network and $v_{j,t}$ be the traffic volume in time interval t at the jth station. Then we define

$$\mathbf{s}_{j,t} = [v_{j,t-p}, \ldots, v_{j,t-1}, v_{j,t}] \tag{1}$$

where p is the order of the time lag. For the target station C, we suppose

$$\begin{cases} \mathbf{x}_{c,t} = \{\mathbf{s}_{j,t} | j = C\} \\ \mathbf{x}_{u,t} = \{\mathbf{s}_{j,t} | j \in \text{upstreams of } C\} \\ \mathbf{x}_{d,t} = \{\mathbf{s}_{j,t} | j \in \text{downstreams of } C\} \\ y_t = \{v_{j,t+1} | j = C\}. \end{cases} \tag{2}$$

Hence, $\mathbf{x}_{c,t}$ is the traffic state vector at current station; $\mathbf{x}_{u,t}$ and $\mathbf{x}_{d,t}$ contain the traffic state vectors from the upstream and downstream stations, respectively. In our work, $\mathbf{x}_{c,t}$, $\mathbf{x}_{u,t}$, and $\mathbf{x}_{d,t}$ constitute the independent variables of the desired prediction model. y_t is the response.

III. Model Description

The MARS model, which was proposed by Friedman [21], is a hybrid nonparametric regression approach that can automatically model nonlinearities and interactions between high-dimensional predictors and responses. MARS has been applied to a wide variety of fields in recent years, including traffic flow prediction [22]. The purpose of this section is to present the theoretical background for Bayesian MARS to prepare for our discussion of its merits and mechanisms when it is applied to the traffic flow prediction problem.

A. Overview of Spatiotemporal MARS

Different from most of the previous work, we feed the traffic states from all of the observation stations into the prediction model and aim at modeling the relationship between all the stations and the target. According to the previous definitions and supposing we have $N + p$ observations at each station, we assume that the response was generated by a model

$$y_t = f(\mathbf{x}_{c,t}, \mathbf{x}_{u,t}, \mathbf{x}_{d,t}) + \epsilon_t, \quad t = 1, 2, \ldots, N \tag{3}$$

where ϵ_t denotes a residual term generated in the stage of traffic data collection, which has zero mean and variance σ^2, and $\epsilon_t \sim N(0, \sigma^2)$. Our aim is to construct an accurate and robust approximation $\hat{f}$ for the function f.

The core idea of MARS is to build a flexible regression function as a sum of basis functions, each of which has its support in a distinct region. Within a region, the regression function reduces to a product of simple functions. In particular, MARS uses the two-sided truncated power basis functions for q-order splines of the form

$$b_q^+(x - \eta) = [+(x - \eta)]_+^q = \begin{cases} (x - \eta)^q, & \text{if } x > \eta \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$b_q^-(x - \eta) = [-(x - \eta)]_+^q = \begin{cases} (\eta - x)^q, & \text{if } x < \eta \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $[\cdot]_+$ is equal to the positive part of the argument, x is the variable split, η is the threshold for the variable, which is named *knot*, and q is the power to which the splines are raised in order to manipulate the degree of the smoothness of the resultant function estimate.

For each predictor $\mathbf{x}_i \in [\mathbf{x}_{c,t}, \mathbf{x}_{u,t}, \mathbf{x}_{d,t}]$, MARS selects the pair of spline functions and the knot location that best describes the response variable. Subsequently, the spline functions are combined into a complex nonlinear model, describing the response as a function of the predictors. Finally, MARS is taken to be a weighted sum of a number of basis functions with the following form:

$$\hat{f}(\mathbf{x}_c, \mathbf{x}_u, \mathbf{x}_d) = \beta_0 + \sum_{m_c=1}^{M_c} \beta_{m_c} B_{m_c}(\mathbf{x}_c) + \sum_{m_u=1}^{M_u} \beta_{m_u} B_{m_u}(\mathbf{x}_u) + \sum_{m_d=1}^{M_d} \beta_{m_d} B_{m_d}(\mathbf{x}_d) \tag{6}$$

where β_0 is a constant bias; β_m are the regression coefficients of the model, which are estimated to yield the best fit to the relationship between the predictor and the response; and $B_m(\mathbf{x})$ is the basis function, or a product of two or more of such functions. In general, the basis functions can be described as the product of L_m univariate spline functions such as

$$B_m(\mathbf{x}) = \prod_{l=1}^{L_m} \phi_{m,l}\left(x_{v(m,l)}\right). \tag{7}$$

Obviously, $B_m(\mathbf{x})$ is the product of L_m univariate spline functions $\{\phi_{m,l}(x_{v(m,l)})\}$, where L_m is the degree of the 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. Thus, for each m, $B_m(\mathbf{x})$ can consist of a single spline function or a product of two or more spline functions, and no input variable can appear more than once in the product. These spline functions are often taken, via (4) and (5), in the following form:

$$\phi_{m,l}\left(x_{v(m,l)}\right) \in \left\{b_q^+\left(x_{v(m,l)} - k_{m,l}\right), b_q^-\left(x_{v(m,l)} - k_{m,l}\right)\right\} \tag{8}$$

where $\eta_{m,l}$ is a knot of $\phi_{m,l}(x_{v(m,l)})$ occurring at one of the observed values of $x_{v(m,l)}$, $l = 1, \ldots, L_m$, $m = 1, \ldots, M$.

When the power of the splines q is equal to 0, the regression function in (6) is equivalent to the regression tree model. Thus, whereas a regression tree model fits a constant at each terminal node, MARS fits more complicated piecewise basis functions in the specific partition.

B. Model Building Using Bayesian Inference

In the MARS model of Friedman, the "optimal" $\hat{f}(\mathbf{x})$ are achieved in a two-stage process: forward growing and backward pruning [21]. However, the method of Friedman only generates one optimal MARS model, which is not stable on large and complex data. Consequently, Denison *et al.* and Holmes and Mallick [23], [24] proposed an MCMC method under the Bayesian inference framework to generate a great number of stable MARS samples. The predicted value is obtained by meaning the responses of these samples. Following these studies, we construct the spatiotemporal MARS model using a Bayesian inference approach. Then, under the defined Bayesian framework, reversible jump MCMC [25] is used to simulate the generation of the MARS sample.

1) Bayesian Inference: When building the model of MARS, the total number of the basis functions M_c, M_u, M_d and the location of the knots expressed via $v(m,l)$ and $\eta_{m,l}$ are the two important factors affecting the accuracy of the MARS model. Being a piecewise model, the number of basis functions in MARS determines the degree of flexibility of the model, and the knots determine the locations of the significant changes in the model.

To find the optimal prediction model, we desire the probability distribution over the space of the possible MARS structure. The candidate structure of the model can be uniquely defined by using the number of basis functions $\{M_c, M_u, M_d\}$, the type of the basis functions $\{B_{m_c}, B_{m_u}, B_{m_d}\}$, and the coefficients $\{\beta_{m_c}, \beta_{m_u}, \beta_{m_d}\}$.

In addition, the type of the basis functions is determined by the degree of the interaction L_m, the index of the variables $v(m,l)$, and the location of the knots $\eta_{m,l}$ according to (8). To find the distribution of possible MARS structure, these arguments are regarded to be random.

Moreover, the Bayesian inference approach places probability distributions on all unknown arguments. Let $\mathcal{M} = \{M_c, M_u, M_d, \mathbf{B}_c, \mathbf{B}_u, \mathbf{B}_d, \boldsymbol{\beta}_c, \boldsymbol{\beta}_u, \boldsymbol{\beta}_d, \sigma^2\}$ refer to a particular model structure and noise variance. Prior distributions on the model space $p(\mathcal{M})$ are updated to posterior distributions by using Bayes' rule, i.e.,

$$p(\mathcal{M}|y) = \frac{p(y|\mathcal{M})p(\mathcal{M})}{p(y)}. \tag{9}$$

Point predictions under the model space can be given as expectations

$$E(y|\mathbf{x}) = \int \hat{f}_{\mathcal{M}}(\mathbf{x}) p(\mathcal{M}|y) d\mathcal{M} \tag{10}$$

where $\mathbf{x} = [\mathbf{x}_c, \mathbf{x}_u, \mathbf{x}_d]$; $\hat{f}_{\mathcal{M}}$ refers to (6) with a set of parameters $\mathcal{M}$.

As the parameter settings $\mathcal{M}$, including the Gaussian error distribution $\epsilon \sim N(0, \sigma^2)$, the marginal log likelihood of the model is expressed by

$$\mathcal{L}(\mathcal{M}|y) = -n \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \left\{ y_i - \hat{f}_{\mathcal{M}}(\mathbf{x}_i) \right\}^2 \tag{11}$$

where n is the number of observations. $\mathcal{L}(\mathcal{M}|y)$ is calculated based on the prior distribution of σ^2 and the coefficients β. In our experiments, the prior distribution of the variance of ϵ is assumed to be following the inverted-gamma (IG) distribution as

$$\sigma^2 \sim \text{IG}(\alpha_1, \alpha_2) \tag{12}$$

where α_1 and α_2 are two parameters controlling the distribution of σ^2. For the coefficients of basis functions, we assume

$$\beta|\sigma^2 \sim N(0, \sigma^2/p_\beta) \tag{13}$$

where p_β is the precision of the coefficient prior.

2) MCMC Simulation: Under the Bayesian framework, our aim is to simulate samples from the posterior distributions $p(\mathcal{M}|y)$. For this purpose, we use the reversible jump MCMC according to the approach of Denison *et al.* [23]. The theory of reversible jump MCMC can be found in [25] for details. In the context of our problem, three options are defined for model-moving strategies.

1) **BIRTH**: Add a basis function, choosing from the temporal, upstream, or downstream predictors uniformly.
2) **DEATH**: Remove one of the existing basis functions uniformly from the present model.
3) **CHANGE**: Change the location of a knot from the model.

In options 1 and 2, the dimension of the model is changed. The probabilities for these three model-moving strategies are assumed to be uniform. After each iteration, the marginal log likelihood of the proposed model and the coefficients are updated. Subsequently, the proposed change to the model is accepted if the exponential of the change of the likelihood is larger than a random value u drawn from the uniform distribution on (0, 1), i.e.,

$$u < \exp\left[\mathcal{L}'(\mathcal{M}'|y) - \mathcal{L}(\mathcal{M}|y)\right] \tag{14}$$

where $\mathcal{M}'$ are the proposed parameters after the model moving, and $\mathcal{L}'$ is the proposed marginal log likelihood. When the number of iterations reaches a predefined number of iterations, the MCMC starts to save the stable samples for the later prediction.

IV. Model Application and Experiments Design

The work in this paper focuses on short-term prediction of the traffic volume on freeways by considering the spatiotemporal correlation property of the traffic flow. Therefore, we employ traffic volume data obtained from observation stations along a long-distance freeway. To verify the capability of our model, the actual traffic data used in the experiments are drawn from the PORTAL FHWA Test Database maintained by Portland State University [26].

A. Data Set Description

The data set used in this paper is collected from eight adjacent stations located on the freeway Interstate 205 (I-205) numbered from South to North. Fig. 1 shows the distribution of the eight chosen observation stations on I-205. There are other two stations on this link. We neglected them in the experiment because there are no traffic data on these two stations. In the figure, the numbers in the circle identify the location of the observation stations.

The traffic volume data were collected between February 24 and March 23, 2013. Univariate traffic volume observations were obtained over intervals of 15 min each. The data collected between February 24 and March 16 are the training data set and divided into weekdays and weekends; this split is also used for evaluating the performance of weekday and weekend prediction models. The traffic volume is formatted as the average number of VPLPH. In Fig. 2, we draw the traffic volume at the eight observation stations on March 18 (Monday).

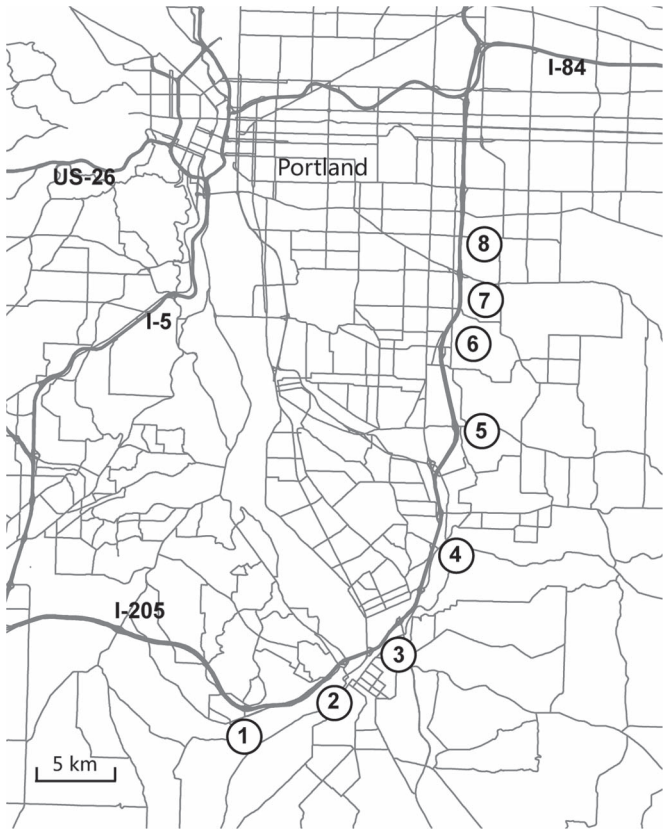

Fig. 1. Locations of the used observation stations on I-205 in Portland.

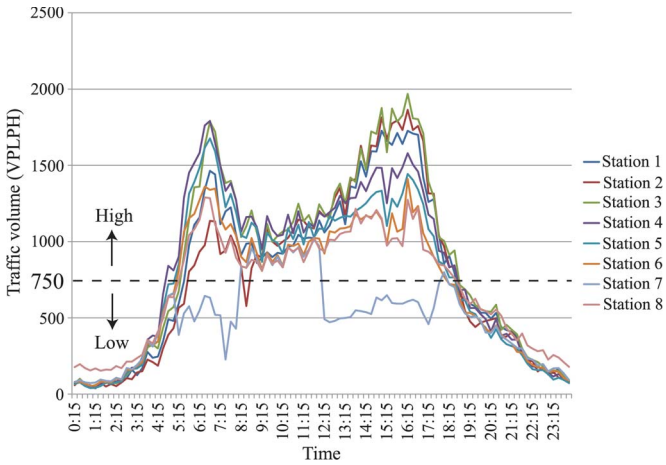

Fig. 2. Traffic volumes at eight stations on March 18 (Monday).

B. Model Application

In the training and testing data sets, the time lag p is set to 3. Therefore, the traffic state vector at interval t for the jth station is $\mathbf{s}_{j,t} = [v_{j,t-3}, v_{j,t-2}, v_{j,t-1}, v_{j,t}]$. If we predict the traffic volume at Station 3, the interrelated variables are defined as $\mathbf{x}_{c,t} = \mathbf{s}_{3,t}$, $\mathbf{x}_{u,t} = [\mathbf{s}_{1,t}, \mathbf{s}_{2,t}]$, $\mathbf{x}_{d,t} = [\mathbf{s}_{4,t}, \ldots, \mathbf{s}_{8,t}]$, and $y_t = v_{3,t+1}$.

In the ST-BMARS model building stage, the order of the basis function q in (4) and (5) is uniformly randomly selected from $\{0, 1\}$. The degree of the interaction of the basis function L_m in (7) is set to 1, that is, the predictors do not interact each other in the basis functions. The index of the predictor composing the basis function $v(m, l)$ is randomly selected from the current, upstream, and downstream state vectors. The location of the knot $\eta_{m,l}$ is randomly selected from $\{1, 2, \ldots, N_{\text{train}}\}$, where N_{train} is the number of observations in the training data set. The maximum sum of basis functions, i.e., $M_{\max} = M_c + M_u + M_d$, is 10.

After defining the type of the desired MARS sample, the main algorithm of MCMC simulation process is given in the Fig. 3. In Fig. 3, the model moving types **BIRTH**, **DEATH**,

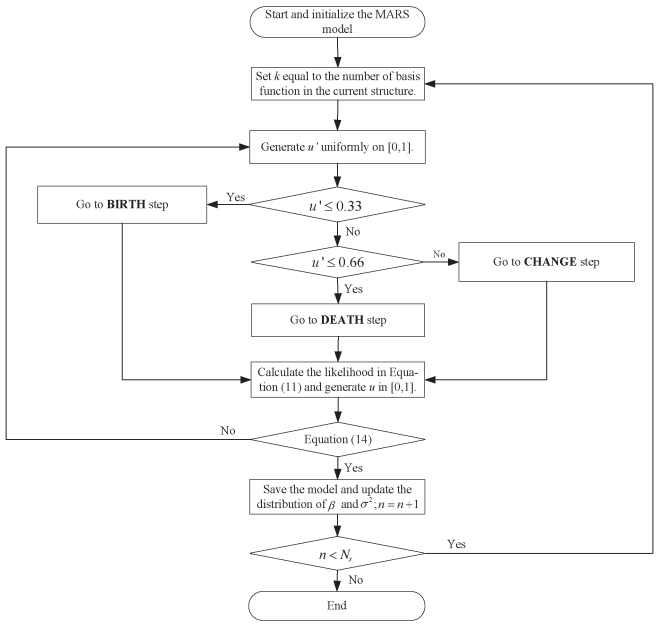

Fig. 3. Main algorithm of the MCMC simulation process.

and **CHANGE** are defined in Section III-B. The pseudocode of **BIRTH** is presented as follows.

BIRTH

1) Uniformly choose the order of the basis function, the position of the knot, the predictor to split on, and the sign indicators in this new basis.
2) Generate u from [0, 1] uniformly.
3) Work out the acceptance probability α.
4) If $(u < \alpha)$, accept the proposed model; else, keep the current model.
5) Return to main algorithm.

The algorithms of **DEATH** and **CHANGE** are similar to **BIRTH**. The parameters are initialized as follows: $n = 0$, $\alpha_1 = \alpha_2 = 0.1$, and $p_\beta = 10$. The maximum number of samples N_s is set to 10 000. When we act the model to the testing data set, the prediction value $\hat{y} = 1/N_s \sum_{n=1}^{N_s} \hat{y}_n$, where $\hat{y}_n$ is the estimation value generated by the nth sample.

C. Comparison Experiments Design

To validate the performance of our proposed prediction model, the temporal MARS (T-MARS) model and three frequently used traffic prediction methods, namely, ARIMA, SARIMA, and SVR, are employed as criterion for comparisons. These models used for comparison are also applied to data sets for weekdays and weekends separately. A brief introduction describes the referenced models.

1) T-MARS: In order to certify the contributions of the spatial traffic states to the object station, a T-MARS model based on historical data is also implemented for comparison. The T-MARS method is implemented using the primordial model proposed by Friedman [21].

2) ARIMA: The ARIMA model is one of the most frequently used parametric techniques in time-series analysis and prediction applications. On the issue of traffic flow prediction, ARIMA is also extensively exploited in practice [6]. In an ARIMA model, the future value of a variable is assumed to be a linear function of several past observations and random errors. We compare the prediction accuracy of our proposed ST-BMARS model with ARIMA since they are both highly interpretable.

In our experiment, ARIMA(3, 0, 1) is employed to predict the traffic volume on the observation stations using their own historical traffic data.

3) SARIMA: The SARIMA model is one of the state-of-the-art parametric techniques and has been successfully applied to the traffic prediction [7], [19], [27]. Through capturing the evident repeating pattern week by week of traffic flow data, the SARIMA introduces weekly dependence relations to the standard ARIMA model and improves the predictive accuracy. In general, the SARIMA model is written as SARIMA$(p, d, q)(P, D, Q)_S$, where p, d, and q are the parameters of the short-term component; P, D, and Q are the parameters related to the seasonal component; and S denotes the seasonal interval.

In our experiment, SARIMA$(1, 0, 1)(0, 1, 1)_S$ is employed to predict the traffic volume on each observation station using their own historical traffic data. For the weekdays model, $S = 96 \times 5 = 480$. For weekends, $S = 96 \times 2 = 192$. To estimate the parameters of the SARIMA model, the model is first represented in state-space form. Next, the parameters are updated using adaptive filtering methods [7]. In our implementation, Kalman filter is used because it can achieve the best predictive accuracy according to the research of Lippi *et al.* [19].

4) SVR: As one of the state-of-the-art nonparametric methods for traffic flow prediction [15], SVR is implemented as a comparison model in this paper. SVR is a kind of kernel function technique based on statistical learning theory developed by Vapnik [28]. It has received increasing attention as a method for solving nonlinear regression problems. SVR is derived from the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error.

In our implementation of the SVR prediction model, we use a radial basis function with parameter $\sigma = 1$ as the kernel function. The SVR model is carried out on the same spatiotemporal information as the proposed ST-BMARS does. Moreover, the best choice of parameters of SVR is determined based on sketching the structure of training data and using a trial-and-error approach.

V. Experiment Results Analysis

In the experiments, we carried the proposed ST-BMARS model on weekdays and weekends, respectively. After obtaining the model, we evaluate the interpretability of the model first. Next, the predictive ability of the ST-BMARS model is compared with other typical prediction models, e.g., ARIMA, SARIMA, and SVR. The robustness of the ST-BMARS model is also analyzed in the last part of this section.

A. Spatiotemporal Relationship Analysis

In traffic engineering practice, when traffic managers design control strategies to alleviate the heavy traffic, the interpretability of the traffic prediction model is particularly important.

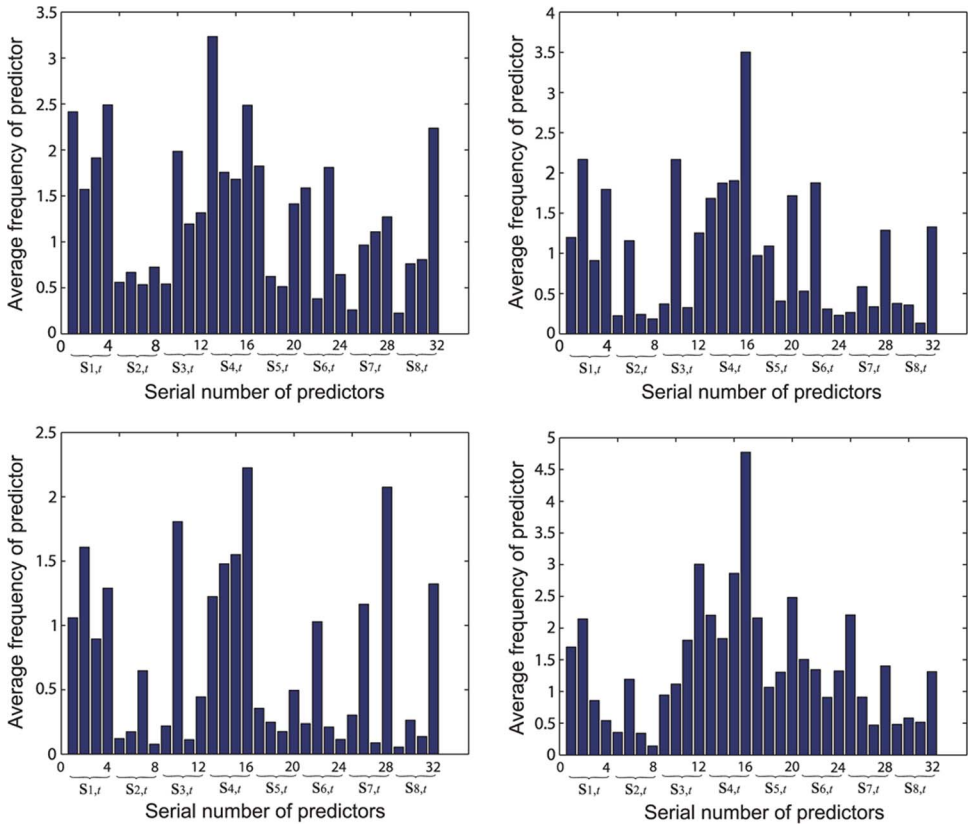

Fig. 4. Importance of predictors for (top to bottom and left to right) Stations 3–6.

An interpretable prediction model can assist traffic managers to make reasonable strategies by extracting the most related stations or road segments that have the greatest contribution to the future traffic state of the target road.

For example, an interpretable model should represent different impacts on future traffic states at a current observation station generated by its historical, upstream, and downstream traffic states. Moreover, the moments when such impacts happen could be also investigated, for instance, steady phase (free flow) or peak time (congestion). Hence, before carrying out the proposed prediction model on the testing data set, the importance of each predictor in the observations to the response volumes is investigated and evaluated first. The contributions of all predictors in (2), including the temporal and spatial information over the eight stations to the current target observation station, are evaluated.

In the traditional MARS model, only one optimal $\hat{f}(\mathbf{x})$ is obtained using the two-stage process of Friedman [21]. Friedman judged the predictor importance via finding reductions of the generalized cross validation after eliminating its basis function from $\hat{f}(\mathbf{x})$. However, in this paper, we generate a great number of MARS samples using MCMC simulation. We track the average frequency of each selected predictor in the samples. We believe that the predictor with high frequency is more important than the one that has low frequency. In other words, the importance of the predictor increases in direct proportion to its frequency in the samples. If a predictor (including spatial and temporal traffic volume) was rarely or never used in any MARS basis function in the samples, we can conclude that it has little or no influence on the specified observation station.

Fig. 4 illustrates the distribution of average frequencies of the predictors over Stations 3–6 in the weekday prediction model. For each station, the set of independent variables $\mathbf{x}$ contains 32 predictors, that is, $\mathbf{x} = [\mathbf{s}_{1,t}, \mathbf{s}_{2,t}, \ldots, \mathbf{s}_{8,t}]$. The values in the horizontal ordinate in Fig. 4 indicate the indices of the predictors in $\mathbf{x}$. The values in the vertical ordinate indicate the average frequencies of the predictors in basis functions of each sample.

The histogram in the upper left in Fig. 4 shows the average frequency of each predictor related to the future traffic volume at Station 3, i.e., $v_{3,t+1}$, on weekdays. From the histogram, we can see that there are five predictors more important than the others. They are $v_{4,t-3}$, $v_{1,t}$, $v_{4,t}$, $v_{1,t-3}$, and $v_{8,t}$, in the order of importance. The most important predictors for Stations 4–6 can be also found in their histograms. For the sake of reflecting the interpretability of the model intuitively, we extract the four most important predictors for each station and plotted them in a relationship graph, as shown in Fig. 5. The width of the line indicates the importance or the contribution of the predictor.

After analyzing the contribution of each predictor to the target variable, we summarize the contribution of each station to the target variable. In this paper, we define the contribution of station as the average of its predictors' contribution (average frequency of predictor), i.e.,

$$\mathcal{C}_{\text{station}} = \frac{1}{p}\sum_{i=1}^{p} z_i \tag{15}$$

where z_i is the average frequency of the ith predictor at the cause station. Then we calculate $\mathcal{C}_{\text{station}}$ for Stations 3–6 and listed the results in Table I.

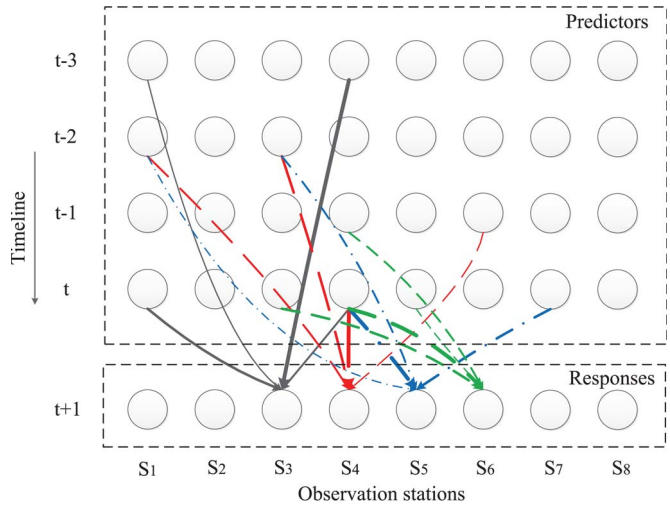

Fig. 5. Most important predictors of the traffic volume at $t+1$ for Stations 3–6; circle denotes the traffic volume.

TABLE I
CONTRIBUTIONS OF EACH STATION TO STATIONS 3–6

$C_{station}(rank)$		Target Station 3	4	5	6
Cause Station	**1**	2.10 *(2)*	1.52 *(2)*	1.21 *(2)*	1.31 *(4)*
	2	0.62 *(8)*	0.45 *(8)*	0.25 *(8)*	0.51 *(8)*
	3	1.26 *(3)*	1.03 *(4)*	0.65 *(4)*	1.72 *(3)*
	4	2.29 *(1)*	2.24 *(1)*	1.62 *(1)*	2.92 *(1)*
	5	1.09 *(5)*	1.04 *(3)*	0.32 *(7)*	1.75 *(2)*
	6	1.11 *(4)*	0.74 *(5)*	0.40 *(6)*	1.27 *(5)*
	7	0.90 *(7)*	0.62 *(6)*	0.91 *(3)*	1.25 *(6)*
	8	1.01 *(6)*	0.55 *(7)*	0.44 *(5)*	0.72 *(7)*

In Figs. 4 and 5 and Table I, we can observe that Stations 4, 1, and 3 generate significant impact to $v_{3,t+1}$. The other stations have comparatively less influence. Similarly, the most significant predictor impacting on Station 4 is its own historical states. This fact illustrates that the future traffic state at Station 4 is more easily influenced by its previous states than other traffic states from upstream or downstream road segments. Moreover, the most significant stations related to Station 5 are Stations 4, 1, and 7. Another particularly noticeable phenomenon is that the historical traffic states at Station 5 have little influence on its future states because the predictors have lower frequencies. As to Station 6, its histogram indicates that Stations 4, 5, and 3 can generate more significant impacts than the other stations.

Furthermore, we also can find that, although Station 2 is at the adjacent upstream of Station 3, it contributes less to $v_{3,t+1}$ than the other stations. A similar phenomenon occurred at Station 7. Station 7 contributes less to $v_{6,t+1}$, although it is at the adjacent downstream of Station 6. We argue that this is due to the different traffic patterns between the adjacent stations. In Fig. 2, we can find that the morning peak around 7:00 at Station 2 is relatively lower, but the evening peak around 16:15 is heavy. The other stations such as Stations 1, 3, 4, and 5 present the two peaks at the same level. Additionally, the traffic pattern at Station 7 is totally different from that at Station 6 in Fig. 2.

Apart from discovering the contributions of the predictors to the future traffic state at the target station, our ST-BMARS model also can interpret how the specific predictor influences the target traffic state. We counted the values of knots η corresponding to the top four most significant predictors to $v_{3,t+1}$ on weekdays in the simulated samples. The histograms of each predictor are illustrated in Fig. 6. From the upper-left histogram, we can see that $v_{4,t-3}$ has a considerable influence on $v_{3,t+1}$ around a high volume. In contrast, when $v_{4,t-3}$ is lower, it generates weak influence on $v_{3,t+1}$. Similarly, we can see from the other histograms that $v_{1,t}$ and $v_{4,t}$ impact on $v_{3,t+1}$ heavily at about 1350 and 1150 VPLPH, respectively. Additionally, the high-frequency knots of $v_{1,t-3}$ are comparatively scattered.

Base on the preceding illustrations and discussions, we can summarize that our model provides the following evident advantages in contrast with previous interpretable parametric models.

1) The interpretability of the proposed model provides spatiotemporal relationship between series stations. Although ARIMA and SARIMA are both interpretable, their predictors are limited to the time domain.
2) The impact weights from each predictor are learned from the traffic data and flexible to different stations. The other spatiotemporal parametric models (STARIMA [20] and MSTAR [13]), by contrast, defined the weights on the basis of the distances between the stations under certain assumptions and constraints.
3) ST-BMARS can quantify the contributions from the observation stations to the future traffic volume at the target station.

B. Prediction Performance Analysis

In practice, the traffic managers are highly concerned with the predictive ability of the system on heavy traffic states. The traffic state, to some extent, can be reflected by the value of the traffic volume. As shown in Fig. 2, the morning and evening peaks are evident at all the stations, except at Station 7. Therefore, to examine the predictive ability of the ST-BMARS model on heavy traffic states, we calculate the prediction errors on the traffic volumes that are larger than 750 VPLPH for both weekdays and weekends. Two measures for prediction error analysis, namely, root-mean-square error (RMSE) and mean absolute percentage error (MAPE), are explored in this research. RMSE and MAPE are defined as

$$\text{RMSE}_{750} = \sqrt{\left[\frac{1}{K}\sum_{k=1}^{K}(V_k - \hat{V}_k)^2\right]} \tag{16}$$

$$\text{MAPE}_{750} = \frac{1}{K}\sum_{k=1}^{K}|V_k - \hat{V}_k|V_k \times 100\% \tag{17}$$

where V_k denotes the actual traffic volume that is larger than 750, during the testing stage, $\hat{V}_k$ is the predicted value produced by the prediction model, and K is the total number of V_k. Additionally, the missing data are not covered in the prediction error evaluation. The values of K when we calculate the prediction errors at each station are listed in Table II.

We discuss the obtained prediction results on weekdays and weekends separately. The averaged values of RMSE and MAPE measures of the involved prediction approaches at each observation station, on five weekdays from March 18 to 22,

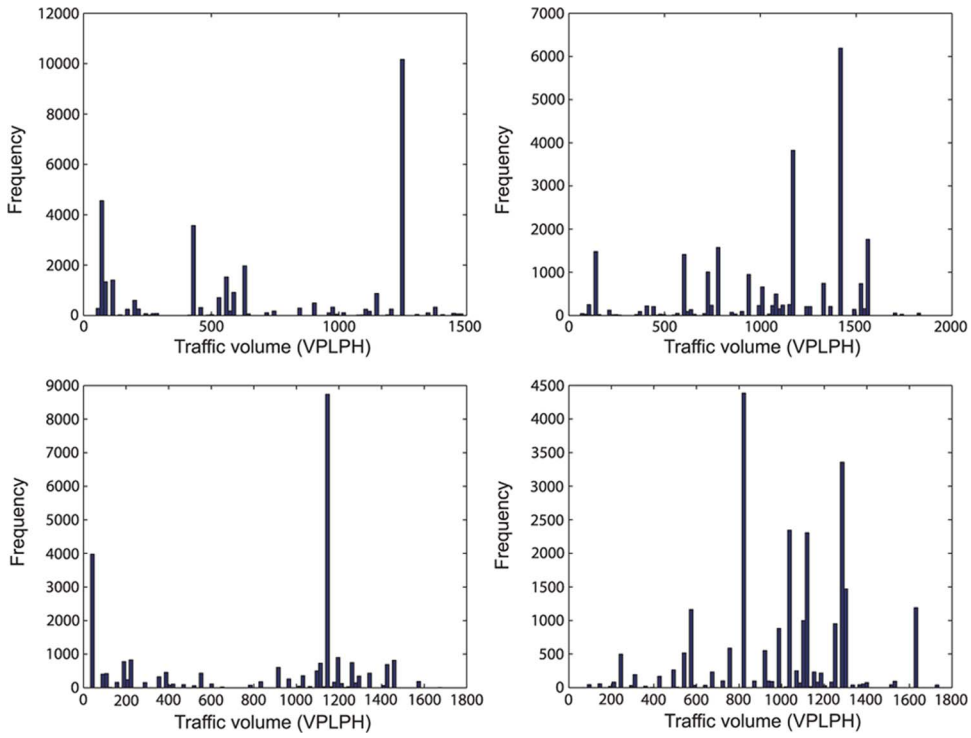

Fig. 6. Histograms of locations of knots over the top four predictors for (top to bottom and left to right) Station 3, $v_{4,t-3}$, $v_{1,t}$, $v_{4,t}$, and $v_{1,t-3}$.

TABLE II
VALUES OF K IN PREDICTION ERROR CALCULATION

	K							
Station	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
weekdays	269	255	279	284	274	258	104	264
weekends	77	74	85	82	77	70	67	74

TABLE III
RMSE OF THE PREDICTION MODELS ON WEEKDAYS

Station	$RMSE_{750}$				
	ST-BMARS	T-MARS	ARIMA	SARIMA	SVR
1	**118.98**	140.53	139.46	149.49	124.20
2	**142.37**	186.94	182.48	152.48	143.20
3	**135.53**	163.16	166.57	171.67	154.63
4	112.03	124.74	129.22	141.05	**110.33**
5	**104.34**	119.02	125.43	137.44	108.37
6	**96.44**	112.02	121.23	122.38	105.58
7	145.08	201.73	168.69	**109.97**	165.46
8	**93.64**	109.97	115.91	126.27	102.09
Average	**118.55**	**144.76**	**143.62**	**138.84**	**126.73**

TABLE IV
MAPE OF THE PREDICTION MODELS ON WEEKDAYS

Station	$MAPE_{750}(\%)$				
	ST-BMARS	T-MARS	ARIMA	SARIMA	SVR
1	**7.36**	8.31	8.45	7.48	7.37
2	**8.51**	10.42	10.52	8.88	9.00
3	**7.64**	9.63	9.73	8.02	8.96
4	6.97	7.57	7.75	7.10	**6.69**
5	**6.80**	7.85	8.07	7.56	7.17
6	**6.88**	7.97	8.50	7.12	7.45
7	10.28	12.09	10.99	**7.58**	11.11
8	**6.65**	7.93	8.16	7.70	7.06
Average	**7.64**	**8.97**	**9.02**	**7.68**	**8.10**

are specified in Tables III and IV, respectively. The RMSE and MAPE measures on weekends (March 17 and 23) are specified in Tables V and VI, respectively.

TABLE V
RMSE OF THE PREDICTION MODELS ON WEEKENDS

Station	$RMSE_{750}$				
	ST-BMARS	T-MARS	ARIMA	SARIMA	SVR
1	**74.48**	90.98	90.23	75.74	84.94
2	111.38	124.44	132.41	**102.00**	141.66
3	65.77	83.70	82.68	**59.14**	84.07
4	72.21	78.87	77.50	**64.75**	71.16
5	61.76	112.28	80.39	**59.32**	70.50
6	**56.87**	66.16	68.08	63.47	63.71
7	**47.60**	59.98	62.87	48.84	57.63
8	**58.73**	85.52	80.96	65.10	70.13
Average	**68.60**	**87.74**	**84.39**	**67.26**	**80.48**

TABLE VI
MAPE OF THE PREDICTION MODELS ON WEEKENDS

Station	$MAPE_{750}(\%)$				
	ST-BMARS	T-MARS	ARIMA	SARIMA	SVR
1	5.92	6.96	6.57	**5.82**	6.32
2	7.76	8.59	8.40	**6.98**	9.49
3	5.07	6.17	6.05	**4.32**	5.99
4	5.22	5.95	5.65	**4.57**	5.21
5	4.72	7.37	6.15	**4.14**	5.20
6	**4.71**	5.54	5.78	5.29	5.40
7	**4.13**	5.16	5.48	4.26	4.78
8	**4.59**	6.46	5.92	5.10	5.29
Average	**5.27**	**6.53**	**6.25**	**5.06**	**5.96**

From Tables III and IV, we can sum up the following conclusions on the prediction errors of the five models on weekdays.

1) The predictive abilities of the T-MARS and ARIMA models are weaker than those of the other three models. These results reflect the fact that these two parametric methods are highly interpretable but always generate doubtable predictions when large quantities of data or nonlinear relationships exist.

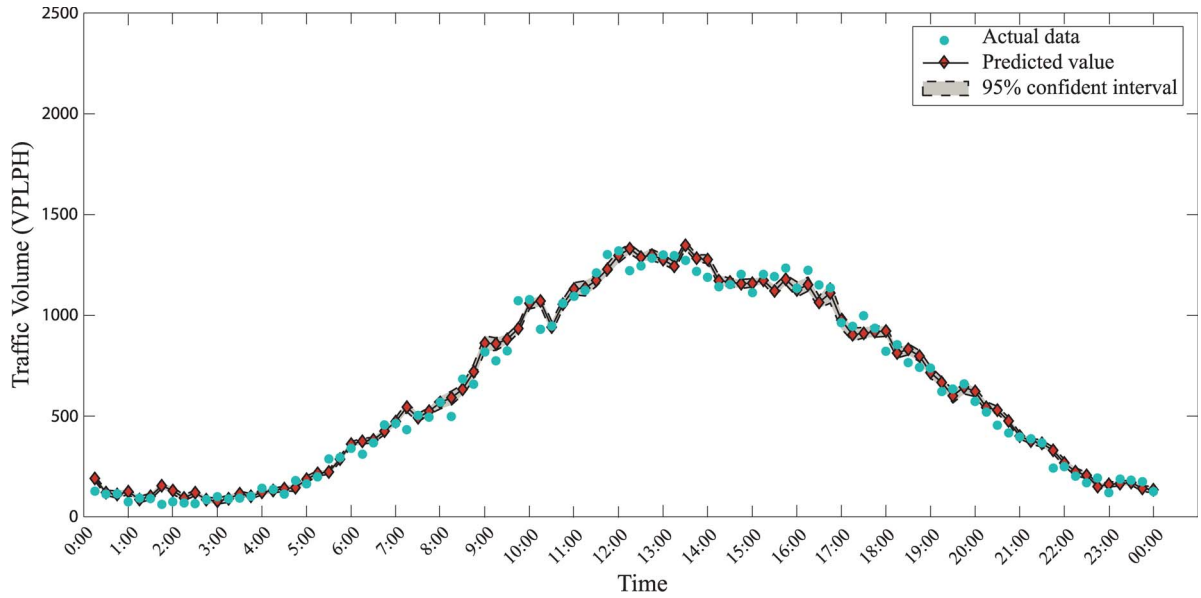

Fig. 7. Prediction results with 95% confidence interval on March 17 (Sunday) at Station 3.

2) The proposed ST-BMARS model performs best at six of eight stations in terms of RMSE_{750} and MAPE_{750}. In particular, compared with the state-of-the-art SARIMA model, ST-BMARS lowers the average RMSE_{750} by 14.6% on the testing weekdays. This promotion indicates that the spatial information could be effectively used to improve the predictive ability of the model.
3) The nonparametric SVR model, which works on the same spatiotemporal information as ST-BMARS, performs better than ST-BMARS only at Station 4. ST-BMARS obtains more accurate prediction than SVR during high traffic volume in terms of MAPE_{750}. This indicates that ST-BMARS utilizes the spatiotemporal information more effectively than SVR.

Furthermore, observing the performances of the prediction models at Station 7, we can find that SARIMA surpasses the other four models, including our ST-BMARS model. That is because, as shown in Fig. 2, the pattern of the traffic volume at Station 7 on weekdays is quite different from those of the other stations in our experiments. In this circumstance, the weekly periodicity of the volume apparently contributes more to the short-term prediction at Station 7 than the spatiotemporal relationship with other stations. If we eliminate Station 7, we can find that the average MAPE_{750} for ST-BMARS and SARIMA are 7.26 and 7.69, respectively.

The prediction errors on the five models on weekends are also specified in Tables V and VI. Owing to the lower complexity of the traffic volume on weekends than weekdays, all of the involved models attain a satisfactory level (MAPE_{750} at all stations are less than 10%). However, the proposed ST-BMARS model still performs best at four stations in terms of RMSE_{750}. SARIMA performs best at the other stations. This indicates that ST-BMARS is competitive at the traffic prediction on weekends.

After our discussion of the five methods in terms of RMSE_{750} and MAPE_{750}, we compare the performances of the involved models at Station 3 in depth. The actual traffic volume and the predicted value by ST-BMARS on March 17 (Sunday) and 18 (Monday) at Station 3 are presented in Figs. 7 and 8, respectively. In the two figures, we also draw the 95% confidence interval of the prediction. As can be seen, the prediction confidence interval is very reliable. That is, ST-BMARS is much confident on its prediction and could predict the actual observation with a lower variance.

The increasing phase of the morning peak and the decreasing phase of the evening peak on March 18 at Station 3 are selected for detailed discussion because they have the steepest slopes within the daily traffic flow, as shown in Fig. 8. The predictions in these periods are unstable due to sudden changes. As shown in Fig. 9, during the increasing phase of the morning peak from 6:15 to 7:00, our ST-BMARS model and the SARIMA model can follow the increase more closely than the other three models. Moreover, ST-BMARS also performs relatively credibly during the decreasing phase from 18:00 to 19:30, as shown in Fig. 10. We also can see that SARIMA fails to predict the traffic volume around 17:30. That is because the prediction of SARIMA is affected by the traffic states at the same time in the last week. If the traffic in the last week was abnormal, the current prediction would be disturbed.

As a consequence of the preceding discussion, the proposed ST-BMARS model improves prediction accuracy on high traffic volume due to incorporating spatial information, as compared with T-MARS. Compared with the highly interpretable ARIMA model, the ST-BMARS model is properly more adaptive to the nonlinear traffic volume. Moreover, the ST-BMARS prediction model outperforms the two state-of-the-art prediction methods, namely, SARIMA and SVR, at most stations, particularly on the heavy traffic on weekdays.

C. Robustness of the ST-BMARS Model

In the model evaluation stage, in addition to the interpretability and accuracy, we also tested the robustness of the proposed

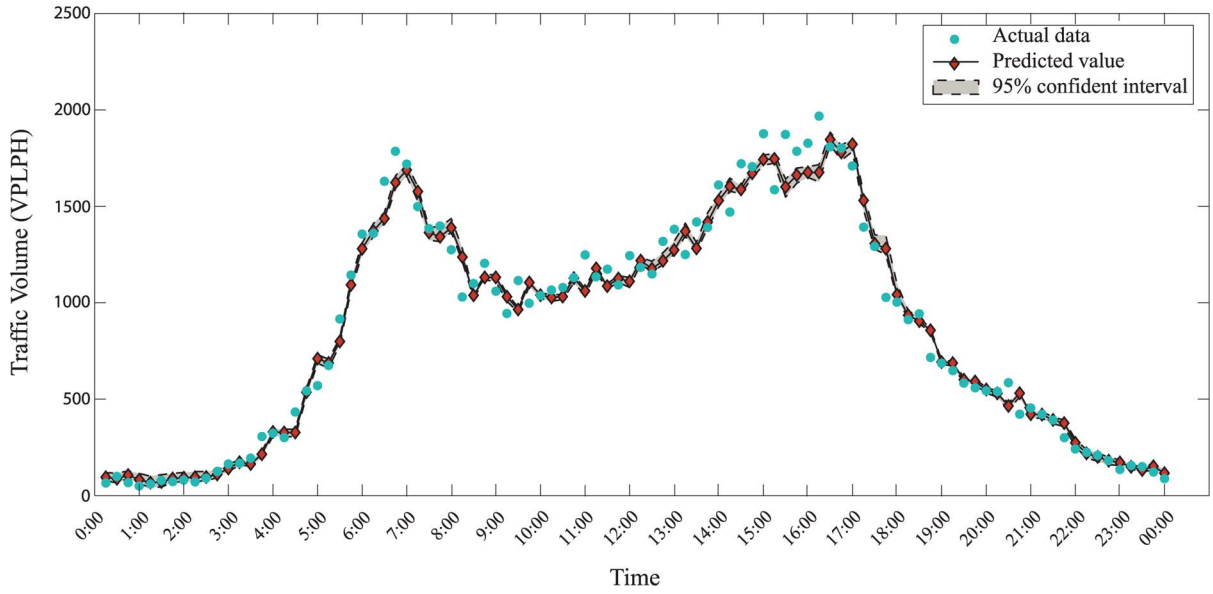

Fig. 8. Prediction results with 95% confidence interval on March 18 (Monday) at Station 3.

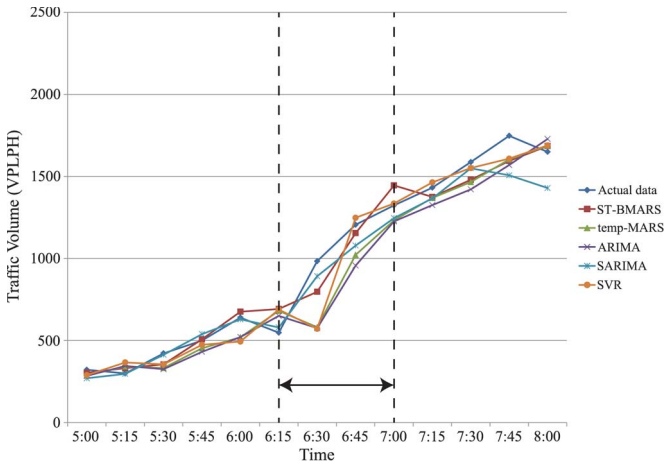

Fig. 9. Increasing phase of traffic on March 18 at Station 3.

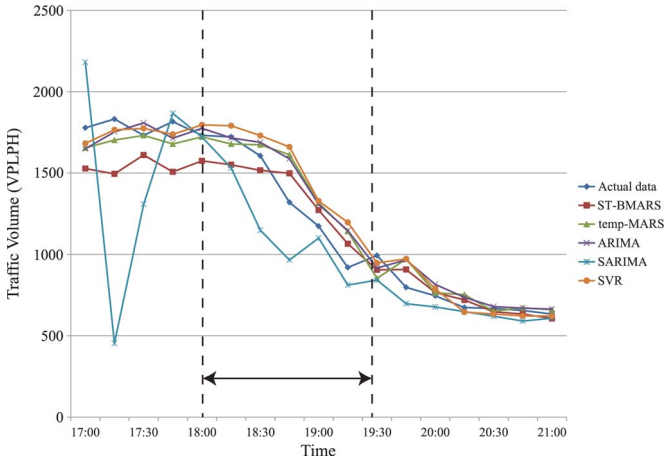

Fig. 10. Decreasing phase of traffic on March 18 at Station 3.

ST-BMARS model. The robustness of the model can be verified from two aspects: robustness to parametric variations and the size of the training data set.

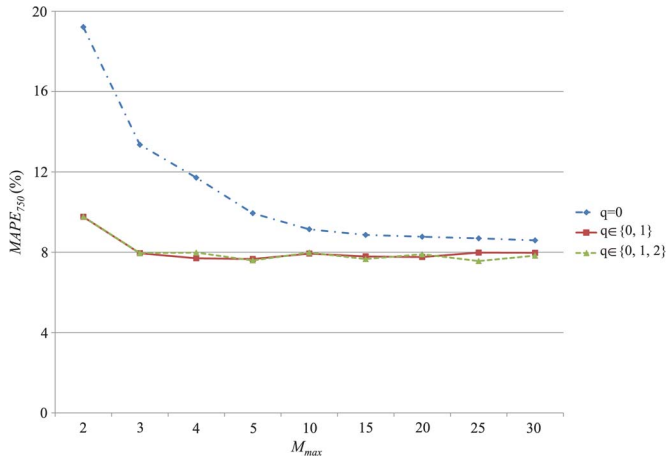

Fig. 11. Robustness testing when the parameters change at Station 3.

In our ST-BMARS model, the two key parameters that control the type of the MARS sample in MCMC simulation are the order of the basis function q and the maximum sum of basis functions $M_{\max}$. In our experiments, the maximum of q was testing from 0 to 2; $M_{\max}$ was selected from 2 to 30. MAPE_{750} is selected as the error criterion. The values of MAPE_{750} at Station 3 on weekdays following the changes of q and $M_{\max}$ are drawn in Fig. 11. As shown in the figure, MAPE_{750} shows a declining tendency with the increase in $M_{\max}$ and becomes stable when $M_{\max} > 5$. Additionally, $q = 0$ generates the worse model than the other two. That is because, when $q = 0$, the MARS model degrades to a regression tree model.

The changes in MAPE_{750} at Station 3 on weekdays with the increase in the size of training data set are drawn in Fig. 12. The X-axis in the figure denotes the number of days used in the model training state, from 3 to 15 days. The figure shows that the error decreases with the increase in the size of training data set. When the number of training days is larger than 13, MAPE_{750} achieves a satisfactory level.

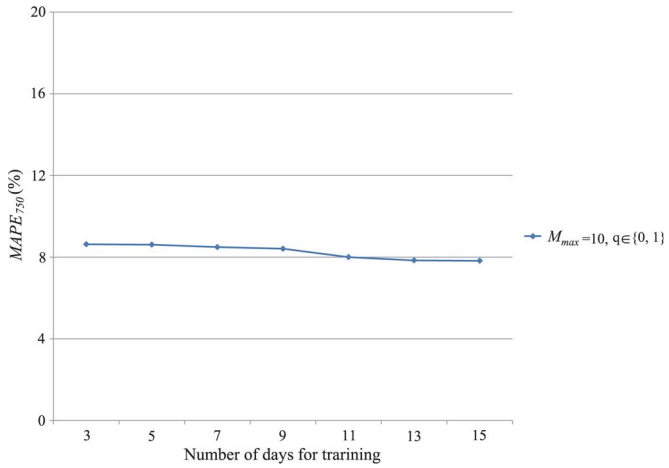

Fig. 12. Robustness testing when the training data set changes at Station 3.

Therefore, based on the preceding illustration, we can conclude that the proposed ST-BMARS model is robust to the variances of model parameters and the size of the training data set.

VI. Conclusion

This paper has proposed an accurate yet interpretable ST-BMARS model for short-term freeway traffic volume prediction. An MCMC simulation was employed to implement the Bayesian inference of the probabilistic model and to obtain a series of stable models. In comparison with previous spatiotemporal correlation models, the proposed model took advantages of the spatial information by selecting significant traffic variables from all of the observation stations along the freeway. The interpretability of the prediction model can assist traffic managers to design reasonable strategies in daily traffic engineering practice.

To verify the effectivity of the ST-BMARS model, experiments were carried out on actual traffic data collected from observation stations on a freeway in Portland, at every 15 min. For comparison, T-MARS, ARIMA, SARIMA, and the kernel SVR method were implemented and compared with the proposed ST-BMARS model in terms of RMSE and MAPE on large volumes. Experimental results indicated that the ST-BMARS model turns out to be a strong contender for short-term freeway traffic volume prediction.

We also notice places that require further improvements for the ST-BMARS model. For example, calculation complexity is high when the model is applied to a large-scale and complex road network. Hence, we are now optimizing the model for large-scale urban traffic networks. Furthermore, we will apply the interpretability of the proposed model to the actual traffic guidance system and evaluate its effectivity in practice.

References

[1] F.-Y. Wang, "Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 630–638, Sep. 2010.

[2] Q.-J. Kong, Y. Xu, S. Lin, D. Wen, F. Zhu, and Y. Liu, "UTN-model-based traffic flow prediction for parallel-transportation management systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1541–1547, Sep. 2013.

[3] Y. Xu, Q.-J. Kong, S. Lin, and Y. Liu, "Urban traffic flow prediction based on road network model," in *Proc. 9th IEEE Int. Conf. Netw., Sens. Control*, Beijing, China, 2012, pp. 334–339.

[4] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: Overview of objectives and methods," *Transp. Rev.*, vol. 24, no. 5, pp. 533–557, Sep. 2004.

[5] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transp. Res. B, Methodol.*, vol. 18, no. 1, pp. 1–11, Feb. 1984.

[6] B. M. Williams, P. K. Durvasula, and D. E. Brown, "Urban freeway travel prediction: Application of seasonal ARIMA and exponential smoothing models," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1644, pp. 132–141, 1998.

[7] S. Shekhar and B. M. Williams, "Adaptive seasonal time series models for forecasting short-term traffic flow," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2024, pp. 116–125, 2007.

[8] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. C, Emerging Technol.*, vol. 10, no. 4, pp. 303–321, Aug. 2002.

[9] T. T. Tchrakian, B. Basu, and M. O'Mahony, "Real-time traffic flow forecasting using spectral analysis," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 519–526, Jun. 2012.

[10] A. Hobeika and C. K. Kim, "Traffic-flow-prediction systems based on upstream traffic," in *Proc. Veh. Navig. Inf. Syst. Conf.*, Yokohama, Japan, 1994, pp. 345–350.

[11] S. Sun, C. Zhang, and G. Yu, "A Bayesian network approach to traffic flow forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 124–132, Mar. 2006.

[12] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 22, no. 5, pp. 317–325, Jul. 2007.

[13] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. C, Emerging Technol.*, vol. 19, no. 4, pp. 606–616, Aug. 2011.

[14] Y. Zhang and Y. Xie, "Forecasting of short-term freeway volume with V-support vector machines," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2024, pp. 92–99, 2007.

[15] M. Castro-Neto, Y.-S. Jeong, M.-K. Jeong, and L. D. Han, "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions," *Exp. Syst. Appl.*, vol. 36, pt. Part 2, no. 3, pp. 6164–6173, Apr. 2009.

[16] S. Sun and X. Xu, "Variational inference for infinite mixtures of Gaussian processes with applications to traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 466–475, Jun. 2011.

[17] Y. Qi and S. Ishak, "Stochastic approach for short-term freeway traffic prediction during peak periods," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 660–672, Jun. 2013.

[18] S. R. Chandra and H. Al-Deek, "Predictions of freeway traffic speeds and volumes using vector autoregressive models," *J. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 53–72, May 2009.

[19] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, Jun. 2013.

[20] Y. Kamarianakis and P. Prastacos, "Space-time modeling of traffic flow," *Comput. Geosci.*, vol. 31, no. 2, pp. 119–133, Mar. 2005.

[21] J. H. Friedman, "Multivariate adaptive regression splines," *Annu. Stat.*, vol. 19, no. 1, pp. 1–67, Mar. 1991.

[22] S. Ye, Y. He, J. Hu, and Z. Zhang, "Short-term traffic flow forecasting based on Mars," in *Proc. 5th Int. Conf. Fuzzy Syst. Knowl. Discov.*, Jinan, China, 2008, vol. 5, pp. 669–675.

[23] D. G. T. Denison, B. K. Mallick, and A. F. M. Smith, "Bayesian MARS," *Stat. Comput.*, vol. 8, no. 4, pp. 337–346, Dec. 1998.

[24] C. C. Holmes and B. K. Mallick, "Bayesian regression with multivariate linear splines," *J. R. Stat. Soc. B, Stat. Methodol.*, vol. 63, no. 1, pp. 3–17, 2001.

[25] P. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, Dec. 1995.

[26] The portal FHWA traffic data set, Portland State University, Portland, OR, USA, Accessed Apr. 11, 2013. [Online]. Available: http://portal.its.pdx.edu/

[27] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.

[28] V. N. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.

Yanyan Xu received the B.E. and M.S. degrees from Shandong University, Shandong, China, in 2007 and 2010, respectively. He is currently working toward the Ph.D. degree in pattern recognition and intelligent systems in the Department of Automation, Shanghai Jiao Tong University, Shanghai, China.

In 2009 he was a Visiting Student with The University of Auckland, Auckland, New Zealand. His research interests include intelligent transportation systems, machine learning, and computer vision.

Qing-Jie Kong (M'07) received the Ph.D. degree in pattern recognition and intelligent systems from Shanghai Jiao Tong University, Shanghai, China, in 2010.

From 2008 to 2009 he was a Visiting Scholar with the Beckman Institute for Advanced Science and Technology, Department of Electrical and Computer Engineering, College of Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA. From 2010 to 2012 he was a Postdoctoral Scientist with the Department of Automation, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University. Since 2012 he has been with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, where he is currently an Associate Professor. His research interests include traffic data mining and fusion, traffic network modeling and analysis, parallel transportation management and control, and video object detection and recognition.

Reinhard Klette received the Ph.D. degree in mathematics, Doctor of Sciences degree and facultas docendi in computer science from Jena University, Germany, in 1978, 1982, and 1984, respectively.

He is a Fellow of the Royal Society of New Zealand and a Professor with The University of Auckland, Auckland, New Zealand. He has coauthored over 300 publications in peer-reviewed journals or conference proceedings and books on computer vision, image processing, geometric algorithms, and panoramic imaging. He has presented over 20 keynotes at international conferences and is the author of *Concise Computer Vision* (London, U.K.: Springer, 2014).

Mr. Klette is the General Chair for the Pacific Rim Symposium on Image and Video Technology 2015 conference at Auckland, New Zealand. He is on the editorial boards of *International Journal of Computer Vision* and *International Journal of Fuzzy Logic and Intelligent Systems*. He was the founding Editor-in-Chief of *Journal of Control Engineering and Technology* in 2011–2013 and an Associate Editor of IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE in 2001–2008.

Yuncai Liu (M'94) received the Ph.D. degree from University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 1990.

From 1990 to 1991 he was an Associate Researcher with Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign. From 1992 to 2000, he was a System Consultant and a Chief Consultant of Research with Sumitomo Electric Industries, Ltd., Tokyo, Japan. In October 2000 he joined Shanghai Jiao Tong University, Shanghai, China, as a Chair Professor of Changjiang Scholarship, Ministry of Education of China, and an Honor Professor. He has authored or coauthored four books and over 200 papers. He is engaged in the wide research fields of computer vision and broad areas in intelligent transportation systems (ITS). He made original contributions in the research studies of 3-D motion estimation, 3-D human motion analysis, and camera calibration. He conducted many advanced projects, such as automatic digital map generation, advanced traffic management systems, advanced traffic information system, vehicle positioning, and vehicle navigation, in the area of ITS. Presently, his research focuses on video detection and cognition, ITS information fusion, and the applications of computer vision in medical surgeries.

Dr. Liu has been a Council Member of the Chinese Transport Engineering Society and the China Society of Image and Graphics, and has been an Associate Editor of *Pattern Recognition*.