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. While in practice, 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 spatio-temporal 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 I-205 in Portland 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 ARIMA, the state-of-the-art seasonal ARIMA model, and the kernel method support vector regression.

Index Terms—Traffic flow prediction, spatio-temporal relationship analysis, interpretable model, MARS, Bayesian inference, MCMC.

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 (ATMS), adaptive traffic-control systems (ATC), or traffic information-services systems (TISS). 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 (SCATS), the Split Cycle Offset Optimization Technique (SCOOT), and parallel-transportation management systems (PtMS) [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 has 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 neighbors (kNN) approach [8], and spectral analysis [9]. These methods can also be 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, in difference to more complicated situations.

In recent years, researchers gradually perceived the significance of spatial information in traffic prediction. Hobekia et al. [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 which 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 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], online support vector regression (SVR) method [15], Gaussian processes [16], a stochastic approach [17], and so forth.

Although the above mentioned spatio-temporal correlation models are quite flexible, they also come with two drawbacks. First of all, most models do not fully exploit the spatial information collected from the whole road network. Previous spatio-temporal 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 especially important. An interpretable model can assist a traffic manager to devise reasonable strategies via extracting the specific road segments which 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 proposed recently 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.

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 State (every 15 minutes) are exploited to verify the effectivity of the proposed prediction model. Afterwards, 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.

There are the following 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 spatio-temporal 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-art prediction methods, such as SARIMA and SVR.

The reminder 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, as well as the robustness of the proposed model are presented and discussed in Section V; 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 that 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. But in traffic engineering practice, when traffic managers design specific strategies to alleviate the heavy traffic, the interpretability of the traffic prediction model is especially important. An interpretable prediction model can assist traffic managers to make reasonable strategies by focusing on the most related stations which 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, especially 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 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.

Afterwards, some interpretable spatio-temporal models are proposed. Kamarianakis et al. [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. But the model is based on the following assumptions: (1) The effect only depends on the distance between the measurement locations; (2) The traffic flow are stable and there are no congestion happened; (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 et al. [13] addressed a multivariate spatio-temporal autoregressive (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. But they only considered such spatial effects from the neighbouring 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 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, as example, loop detector can obtain the number of passing vehicles, the occupancy and the speed. But the occupancy and 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 especially 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 supports 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

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

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

where \( [.]_+ \) equals the positive part of the argument; \( x \) is variable split and \( \eta \) is the threshold for the variable, named knot; \( q \) 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 \( x_i \in [x_{c,t}, x_{u,t}, x_{d,t}] \), 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:

$$\hat{f}(x_c, x_u, x_d) = \beta_0 + \sum_{m_c=1}^{M_c} \beta_{mc} B_{mc}(x_c)$$

$$+ \sum_{m_u=1}^{M_u} \beta_{mu} B_{mu}(x_u) + \sum_{m_d=1}^{M_d} \beta_{md} B_{md}(x_d)$$

where \( \beta_0 \) is a constant bias; \( \beta_m \) are the regression coefficients of the model, which are estimated to yield the best fit to the

III. MODEL DESCRIPTION

The MARS model, proposed by Friedman [21], is a hybrid nonparametric regression approach which 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 Spatio-Temporal 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 the response was generated by a model

$$y_t = f(x_{c,t}, x_{u,t}, x_{d,t}) + \epsilon_t, \quad t = 1, 2, \ldots, N$$

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

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

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

where \( [.]_+ \) equals the positive part of the argument; \( x \) is variable split and \( \eta \) is the threshold for the variable, named knot; \( q \) 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 \( x_i \in [x_{c,t}, x_{u,t}, x_{d,t}] \), 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:

$$\hat{f}(x_c, x_u, x_d) = \beta_0 + \sum_{m_c=1}^{M_c} \beta_{mc} B_{mc}(x_c)$$

$$+ \sum_{m_u=1}^{M_u} \beta_{mu} B_{mu}(x_u) + \sum_{m_d=1}^{M_d} \beta_{md} B_{md}(x_d)$$

where \( \beta_0 \) is a constant bias; \( \beta_m \) are the regression coefficients of the model, which are estimated to yield the best fit to the
relationship between the predictor and response; $B_m(x)$ 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 as

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

Obviously, $B_m(x)$ 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 $m$th basis function and the $l$th spline function. Thus, for each $m$, $B_m(x)$ 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 in the following form, via Equations (4) and (5),

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

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 Equation (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 Friedman’s MARS model, the “optimal” $\hat{f}(x)$ are achieved in a two-stage process: forward growing and backward pruning [21]. However, Friedman’s method only generates one “optimal” MARS model, which is not stable on large and complex data. Consequently, Denison et al. [23], [24] proposed a MCMC method under the Bayesian inference framework to generate a great number of stable MARS samples. The predicted value is obtained by using the responses of these samples. Following these studies, we construct the spatio-temporal 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.

For the purpose of finding 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 are 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 equation (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 $\mathcal{M} = \{M_c, M_u, M_d, B_{m_c}, B_{m_u}, B_{m_d}, \beta_{m_c}, \beta_{m_u}, \beta_{m_d}, \sigma^2\}$ 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

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

Point predictions under the model space can be given as expectations

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

where $x = [x_c, x_u, x_d]$; $\hat{f}(x)$ refers to Equation (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

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

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

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

where $\alpha_1$ and $\alpha_2$ are two parameters controlling the distribution of $\sigma^2$. For the coefficients of basis functions we assume

$$\beta_0 \sigma^2 \sim N(0, \sigma^2/p_\beta) \tag{13}$$

where $p_\beta$ 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 using the reversible jump MCMC according to Denison et al.’s approach [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:

(a) **BIRTH**: add a basis function, choosing from the temporal, upstream, or downstream predictors uniformly;
(b) **DEATH**: remove one of the existing basis functions uniformly from the present model;
(c) **CHANGE**: change the location of a knot from the model.

In options (a) and (b), 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 log-likelihood is larger than a random value $u$ drawn from the uniform distribution on $(0, 1)$.

$$u < \exp[\mathcal{L}(\mathcal{M} | y) - \mathcal{L}(\mathcal{M} | y)] \tag{14}$$
where $\mathcal{M}'$ is the proposed parameters after the model moving; $\mathcal{L}'$ 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 spatio-temporal 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 is 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. Figure 1 shows the distribution of the eight chosen observation stations on the I-205. Actually, 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 to March 16. Univariate traffic volume observations were obtained every 15 minutes each. The data collected between February 24 to March 16 is 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 vehicles per lane per hour (VPLPH). In Figure 2, we draw the traffic volume at the eight observation stations on 18 March (Monday).

B. Model Application

In the training and testing data sets, the timelag $p$ is set to 3. Therefore, the traffic state vector at interval $t$ for the $j$th station is $s_{j,t} = [v_{j,t-3}, v_{j,t-2}, v_{j,t-1}, v_{j,t}]$. If we predict the traffic volume at Station 3, the interrelated variables are defined as follows: $x_{c,t} = s_{3,t}$, $x_{u,t} = [s_{1,t}, s_{2,t}]$, $x_{d,t} = [s_{4,t}, \ldots, s_{8,t}]$, and $y_{t} = v_{3,t+1}$.

In the ST-BMARS model building stage, the order of the basis function $q$ in Equation (4, 5) is uniformly randomly selected from $\{0, 1\}$. The degree of the interaction of the basis function $L_m$ in Equation (7) is set to 1, that is, the predictors don’t 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 $N_{\text{train}}$ is the number of observations in the training data set. The maximum sum of basis functions, $M_{\text{max}} = M_c + M_u + M_d$, is 10.

After defining the type of desired MARS sample, the main algorithm of MCMC simulation process is given in the Figure 3:

In Figure 3, the model moving types BIRTH, DEATH, and CHANGE are defined in section III-B. The pseudo-code 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, $\alpha$.
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: $n = 0$, $\alpha_1 = \alpha_2 = 0.1$, $p_\beta = 10$. The maximum number of samples $N_s$ is set to 10000. When we act the model to the testing data...
set, the prediction value \( \hat{y} = 1/N \sum_{n=1}^{N_s} \hat{y}_n \), where \( \hat{y}_n \) is the estimation value generated by the \( n \)th sample.

C. Comparison Experiments Design

To validate the performance of our proposed prediction model, the temporal MARS model as well as three frequently used traffic prediction methods, 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) Temporal MARS: In order to certify the contributions of the spatial traffic states to the object station, a temporal MARS model (T-MARS) based on historical data is also implemented for comparison. The temporal 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 both are 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) Seasonal ARIMA: 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 \( p, d, q \) are the parameters of the short-term component, \( P, D, Q \) are the parameters related to the seasonal component, \( S \) 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 Lippi et al.’s research [19].

4) Support Vector Regression: As one of the state-of-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 et al. [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 (RBF) with parameter \( \sigma = 1 \) as the kernel function. The SVR model is carried out on the same spatio-temporal information as the proposed ST-BMARS does. Moreover, the best choice of parameters of the SVR are 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 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. Spatio-Temporal 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 especially important. An interpretable prediction model can assist traffic managers to make reasonable strategies by extracting the most related stations or road segments which 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 also be investigated, steady phase (free flow) or peak time (congestion), for instance. 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 Equation (2), including the temporal and the spatial information over the eight stations to the current target observation station, are evaluated.
In the traditional MARS model, only one “optimal” \( \hat{f}(x) \) is obtained using Friedman’s two-stage process [21]. Friedman judged the predictor importance via finding reductions of the generalized cross validation (GCV) after eliminating its basis function from \( \hat{f}(x) \). 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.

Figure 4 illustrates the distribution of average frequencies of the predictors over the Stations 3, 4, 5, and 6 in the weekday prediction model. For each station, the set of independent variables \( x \) contains 32 predictors, that is \( x = [s_{1,t}, s_{2,t}, \ldots, s_{8,t}] \). The values in the horizontal ordinate in Figure 4 indicate the indices of the predictors in \( x \). 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 of Figure 4 shows the average frequency of each predictor related to the future traffic volume at Station 3, \( v_{3,t+1} \), on weekdays. From the histogram, we can see that there are five predictors are more important than 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, 5, and 6 can also be 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 in a relationship graph, as shown in Figure 5. The width of the line indicates the importance or 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).

\[
C_{\text{station}} = \frac{1}{p} \sum_{i=1}^{p} z_i
\]  

where \( z_i \) is the average frequency of the \( i \)th predictor at the cause station. Then we calculate \( C_{\text{station}} \) for Stations 3, 4, 5, and 6, and listed the results in Table I.

### Table I

<table>
<thead>
<tr>
<th>( C_{\text{station}}(\text{rank}) )</th>
<th>Target Station</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>2.10 (2)</td>
</tr>
<tr>
<td>2</td>
<td>0.62 (8)</td>
</tr>
<tr>
<td>3</td>
<td>1.26 (3)</td>
</tr>
<tr>
<td>4</td>
<td>2.29 (1)</td>
</tr>
<tr>
<td>5</td>
<td>1.09 (5)</td>
</tr>
<tr>
<td>6</td>
<td>1.11 (4)</td>
</tr>
<tr>
<td>7</td>
<td>0.90 (7)</td>
</tr>
<tr>
<td>8</td>
<td>1.01 (6)</td>
</tr>
</tbody>
</table>

From Figures 4, 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 other stations.

Furthermore, we also can find that although Station 2 is at the adjacent upstream of Station 3, it contributions less to
In practice, the traffic managers are highly concerned with
the predictive ability of the system on heavy traffic states. The
target traffic state, to some extent, can be reflected by the value
of the traffic volume. As shown in Figure 2, the morning and
evening peak are evident at all the stations except 7. Therefore,
to examine the predictive ability of the ST-BMARS model on
a heavy traffic states, we calculate the prediction errors on
the traffic volumes which are larger than 750 VPLPH for both
weekdays and weekends. Two measures for prediction error
analysis, root mean square error (RMSE) and mean absolute
percentage error (MAPE) are explored in this research. RMSE
and MAPE are defined as follows:

\[
RMSE_{750} = \sqrt{\frac{1}{K} \sum_{k=1}^{K} (V_k - \hat{V}_k)^2}
\]

\[
MAPE_{750} = \frac{1}{K} \sum_{k=1}^{K} \left| \frac{V_k - \hat{V}_k}{V_k} \right| \times 100\%
\]

where \(V_k\) denotes the actual traffic volume which is larger
than 750 during the testing stage; \(\hat{V}_k\) is the predicted value
produced by the prediction model; \(K\) is the total number of
\(\hat{V}_k\). 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.

<table>
<thead>
<tr>
<th>Station</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>weekdays</td>
<td>269</td>
<td>255</td>
<td>279</td>
<td>284</td>
<td>274</td>
<td>258</td>
<td>104</td>
<td>264</td>
</tr>
<tr>
<td>weekends</td>
<td>77</td>
<td>74</td>
<td>85</td>
<td>82</td>
<td>77</td>
<td>70</td>
<td>67</td>
<td>74</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>RMSE of the prediction models on weekdays</th>
</tr>
</thead>
<tbody>
<tr>
<td>Station</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
</tbody>
</table>

Average | 118.55 | 144.76 | 143.62 | 138.84 | 126.73 |
Fig. 7. Prediction results with 95% confident interval on March 17 (Sunday) at Station 3

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

TABLE IV

<table>
<thead>
<tr>
<th>Station</th>
<th>MAPE_{750} (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7.36</td>
</tr>
<tr>
<td>2</td>
<td>8.51</td>
</tr>
<tr>
<td>3</td>
<td>7.64</td>
</tr>
<tr>
<td>4</td>
<td>6.91</td>
</tr>
<tr>
<td>5</td>
<td>6.80</td>
</tr>
<tr>
<td>6</td>
<td>6.88</td>
</tr>
<tr>
<td>7</td>
<td>10.28</td>
</tr>
<tr>
<td>8</td>
<td>6.65</td>
</tr>
<tr>
<td>Average</td>
<td>7.64</td>
</tr>
</tbody>
</table>

TABLE V

<table>
<thead>
<tr>
<th>Station</th>
<th>RMSE_{750}</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>74.48</td>
</tr>
<tr>
<td>2</td>
<td>111.38</td>
</tr>
<tr>
<td>3</td>
<td>65.77</td>
</tr>
<tr>
<td>4</td>
<td>72.21</td>
</tr>
<tr>
<td>5</td>
<td>61.76</td>
</tr>
<tr>
<td>6</td>
<td>56.87</td>
</tr>
<tr>
<td>7</td>
<td>47.60</td>
</tr>
<tr>
<td>8</td>
<td>58.73</td>
</tr>
<tr>
<td>Average</td>
<td>68.60</td>
</tr>
</tbody>
</table>
From Table 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 model 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 doubtful predictions when large quantities of data or nonlinear relationships exist.

2. The proposed ST-BMARS model performs best at six of eight stations in terms of $RMSE_{750}$ and $MAPE_{750}$. Especially 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 spatio-temporal information as ST-BMARS, performs better than ST-BMARS only at Station 4. The ST-BMARS obtains more accurate prediction than SVR during high traffic volume in term of $MAPE_{750}$. This indicates that ST-BMARS utilizes the spatio-temporal information more effectively than SVR.

Furthermore, observing the performances of the prediction models at Station 7, we can find that the SARIMA surpasses the other four models including our ST-BMARS model. That’s because, as shown in Figure 2, the pattern of the traffic volume at Station 7 on weekdays is quite different from 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 spatio-temporal relationship with other stations. If we eliminate Station 7, we can find that the average $MAPE_{750}$ for ST-BMARS and ARIMA are 7.26 and 7.69, respectively.

The prediction errors on the five models on weekends are also specified in Table 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 still performs best at four stations in term of $RMSE_{750}$. The SARIMA performs best at the other stations. This indicates that ST-BMARS is competitive at the traffic prediction on weekends.

After our discussion of 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 Figure 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, the ST-BMARS is much confident to 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 Figure 8. The predictions in these periods are unstable due to sudden changes. As shown in Figure 9, during the increasing phase of the morning peak from 6:15 to 7:00, our ST-BMARS and the SARIMA model can follow the increase more closely than the other three models. Moreover, the ST-BMARS also performs sufficiently credibly during the decreasing phase from 18:00 to 19:30 as shown in Figure 10. We also can see that the SARIMA fails to predict the traffic volume around 17:30. That’s because the prediction of SARIMA is affected by the traffic statues at the same time in last week. If the traffic in last week was abnormal, the current prediction would be disturbed.

In consequence of the above discussion, the proposed ST-
BMARS improves prediction accuracy on high traffic volume due to incorporating spatial information compared to the temporal MARS. Compared to 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, SARIMA and SVR, at most stations, especially on the heavy traffic on weekdays.

C. Robustness of the ST-BMARS model

In the model evaluation stage, besides the interpretability and accuracy, we also tested the robustness of the proposed 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.

In our ST-BMARS, the two key parameters 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. The \( MAPE_{750} \) 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 Figure 11. As the figure shows, \( MAPE_{750} \) shows a declining tendency with the increase of \( M_{max} \) and becomes stable when \( M_{max} > 5 \). Additionally, \( q = 0 \) generates the worse model than the other two. That’s because when \( q = 0 \), the MARS model degrades to regression tree model.

The changes of \( MAPE_{750} \) at Station 3 on weekdays with the increase of the size of training data set are drawn in Figure 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 of the size of training data set. When the number of training days larger than 13, the \( MAPE_{750} \) achieves a satisfactory level.

Therefore, based on the above illustrates, 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. CONCLUSIONS

This paper proposed an accurate yet interpretable spatio-temporal Bayesian multivariate adaptive-regression splines (ST-BMARS) model for short-term freeway traffic volume prediction. A Markov-chain Monte Carlo (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 spatio-temporal 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 minutes. For comparison, the temporal MARS, the ARIMA, the seasonal ARIMA, 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 which 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


[24] Yanyan Xu received the B.E. and M.S. degree from the School of Information Science and Engineering, Shandong University, China, in 2007 and 2010, respectively. In 2009 she was a visiting student at the University of Auckland, New Zealand. He is currently working towards a Ph.D. degree in pattern recognition and intelligent systems, at the Department of Automation, Shanghai Jiao Tong University, China.

His research interests include intelligent transportation systems, machine learning, and computer vision.

[25] Qing-Jie Kong (S’07–M’10) 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 Ph.D. student in the Beckman Institute for Advanced Science and Technology, the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

He was a Postdoctoral Researcher at the Department of Automation, Shanghai Jiao Tong University, from 2010 to 2012. Since 2012, he is Assistant Professor at the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His research interests include traffic data mining and fusion, traffic network modeling and analysis, and video object detection and recognition.

[26] Reinhard Klette is a Fellow of the Royal Society of New Zealand and a professor at the University of Auckland, Auckland, New Zealand.

He is currently the Editor-in-Chief of the Journal of Control Engineering and Technology, on the editorial boards of the International Journal of Computer Vision and the International Journal of Fuzzy Logic and Intelligent Systems, and was an associate editor of IEEE PAMI in 2001–2008. He edited over 100 publications in peer-reviewed journals or conferences, and books on computer vision, image processing, geometric algorithms, and panoramic imaging. He presented more than 20 keynote speeches at international conferences.

[27] Yuncai Liu received a Ph.D. degree from the University of Illinois at Urbana-Champaign, in the Department of Electrical and Computer Engineering in 1990.

He was an Associate Researcher in the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, from 1990 to 1991. Since 1992, he has been a System Consultant and a Chief Consultant of Research in Sumitomo Electric Industries, Ltd., Japan. In October 2000, he joined the Shanghai Jiao Tong University as a Chair Professor of Changjiang Scholarship of the Ministry of Education and Honor Professor of Shanghai Jiao Tong University. He has published 4 books and more than 200 papers. He is engaged in the wide research fields of computer vision and broad areas in Intelligent Transportation Systems. He made original contributions in the researches of 3D motion estimation, 3D human motion analysis and camera calibration. In the area of Intelligent Transportation Systems, he conducted many advanced projects, such as automatic digital map generation, advanced traffic management systems, advanced traffic information system, vehicle positioning, and vehicle navigation. Presently, his research performances focus on video detection and cognition, ITS information fusion and the applications of computer vision in medical surgeries.

Prof. Liu has been an associate editor of Pattern Recognition, a council member of Chinese Transport Engineering Society, and a council member of China Society of Image and Graphics.