Short-Term Traffic State Prediction Based on Temporal–Spatial Correlation

T. L. Pan, A. Sumalee, R. X. Zhong, and N. Indra-payoong

***Abstract*—The stochastic cell transmission model (SCTM) was originally developed for stochastic dynamic traffic state modeling under several assumptions, e.g., the independent/uncorrelated assumption of the underlying stochastic processes governing demand and supply uncertainties. However, traffic flow, by nature, is correlated in both spatial and temporal domains due to its dynamics, similar environmental conditions and human behaviors. The independent assumption in the original SCTM framework may prevent the model from a broad range of applications, e.g., short-term traffic state prediction. In this paper, the SCTM framework is extended to consider the spatial–temporal correlation of traffic flow and to support short-term traffic state prediction. First, a multivariate normal distribution (MND)-based best linear predictor is adopted as an auxiliary dynamical system to the original SCTM to forecast boundary variables and/or supply functions. The predicted boundary variables and supply functions are taken as inputs to the SCTM to perform short-term traffic state prediction. The independent assumption of the SCTM is relaxed by incorporating the covariance structure calibrated from the spatial correlation analysis for probabilistic traffic state evaluation. For real-time application purposes, prediction is conducted in a rolling horizon manner, which is useful for adjusting the predicted traffic state using real-time measurements. The proposed traffic state prediction framework is validated by empirical studies that demonstrate the effectiveness of the proposed method.**

***Index Terms*—Abnormal traffic conditions, rolling horizon, spatial–temporal correlation, stochastic cell transmission model (SCTM), traffic state prediction.**

I. Introduction

TRAFFIC networks are exposed to both demand and supply uncertainties. The SCTM extended the cell transmission model (CTM) to consider the effects of supply uncertainty (i.e., stochastic parameters of the fundamental flow–density diagram) and demand uncertainty (boundary variables of a freeway corridor are regarded as demand) on the distribution of traffic flow for a freeway corridor [20]. All noise terms governing the stochastic supply and demand are assumed to be described by some wide-sense stationary second-order processes, which are uncorrelated in both space and time domains. Due to similar environmental conditions, human behaviors, high traffic density, congestion in the network, and the interaction of demand and supply uncertainties along with the dynamic nature of traffic flow, the demand and supply uncertainties are correlated in both space and time domains. For example, free-flow speeds are spatially correlated (cell-to-cell, lane-to-lane correlated) [11]; demand profiles are temporally correlated [31]; and the traffic state of a specific site is highly affected by its upstream and downstream traffic conditions [5], [15], [18]. In this paper, spatial correlations refer to calibrated correlations of supply functions at different locations, e.g., correlations of free-flow speeds along a freeway corridor, whereas temporal correlations refer to the dependence of traffic variables in the time domain. Considering spatial and temporal correlations along with traffic dynamics bring significant potential advantages for the development of efficient traffic state prediction for the SCTM paradigm.

Short-term traffic prediction aims at evaluating anticipated traffic conditions at a future time given the historical traffic information in the "near future," which has a horizon of only a few minutes, and real-time detected traffic information [23], [27]. Short-term traffic state prediction is one of the critical components of advanced travelers' information systems (ATIS) and advanced traffic management systems (ATMS) [12]. Major techniques for short-term traffic state prediction can be divided into three categories in transportation literature, i.e., 1) parametric methods (see, e.g., [10] and [27] for an overview), such as dynamic-traffic-assignment-based models (see, e.g., [17] for an overview) and Kalman filtering techniques; 2) nonparametric statistical methods, e.g., neural network models, k-nearest neighbor algorithm (k-NN) [23], simulation models, Bayesian models, and support vector regression (see [10] for an overview); and 3) hybrid integration methods that combine the parametric and nonparametric methods.

Some researchers may group these methods into statistical-based and computational intelligence (CI)-based methods. The statistical approaches have more solid and widely accepted mathematical foundations than the CI-based approaches. However, none of these classifications are persuasive enough to gain the consensus of all researchers. The statistical approaches, by their very nature the mathematics of collecting,

Manuscript received October 8, 2012; revised February 8, 2013; accepted April 2, 2013. Date of publication May 17, 2013; date of current version August 28, 2013. This work was supported in part by the Hong Kong Research Grants Council General Research Fund under Grant PolyU 5250/11E, by the Faculty of Construction and Environment Postdoctoral Fellowship Scheme under Project 1-ZV9L, and by the National High-Tech R&D Program of China (863 Program) under Grant 2012AA112311. The Associate Editor for this paper was Y. Wang.

T. L. Pan and A. Sumalee are with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong (e-mail: ceasumal@polyu.edu.hk).

R. X. Zhong is with the Research Center of Intelligent Transportation Systems, School of Engineering, Sun Yat-sen University, Guangzhou, China, and also with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong.

N. Indra-payoong is with the Department of Logistics and Supply Chain Management, Burapha University, Bangkok 20131, Thailand.

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.2013.2258916

organizing, and interpreting numerical data, can also provide more insights on the mechanisms creating and processing the data, particularly when the data concern the analysis of population characteristics by inference from sampling [10]. On the other hand, the statistical approaches frequently fail when dealing with complex and highly nonlinear data and suffer from the curse of dimensionality, e.g., the underlying assumptions of Kalman filtering are: the system is observable; the system has no strong nonlinearity (so that the system dynamics can be well approximated by its linearization); and the model parameters[1] are precisely known [1], [9], [27]. On the other hand, an autoregressive integrated moving-average (including seasonal autoregressive integrated moving average) approach is highly computationally intensive, even on only two sensor locations [15].

The similarities and differences of the statistical- and CI-based approaches have been discussed in a recent overview paper [10] and technical papers [1]–[4], [19], [24]. In general, it can be inferred from the review that, for the time being, no single method can consistently outperform the others. Moreover, comparison works showed that those complex theoretically sound methods, e.g., complex statistic models and neural networks, might not be superior to the simpler methods, e.g., historical methods. (We refer the readers to [10] and [27] for detailed comparison and discussion of these approaches.) However, we may point out that abnormal traffic patterns caused by nonrecurrent congestion or incidents may deteriorate the performance of these models (see, e.g., [25] and [28]). Nevertheless, under most situations, extreme values are of primary interest in forecasting the change in traffic conditions.

Most of the existing short-term traffic state prediction methods are based on the autocorrelation functions of the traffic variables at a specific location, e.g., the location with measurement devices [27], and do not consider traffic flow theory. Under this circumstance, on the one hand, traffic state information on upstream and downstream locations has been largely neglected in these traffic prediction methods. On the other hand, traffic dynamics such as shockwaves, queue formation, and queue dissipation and dynamic traffic interactions across multiple links such as queue spillback cannot be well captured [22]. To overcome this, hybrid approaches that combine the complementary features and capabilities of both traffic flow models, e.g., the CTM in [22] and the Lighthill–Witham–Richards model in [13], and knowledge-based models, e.g., statistical models, were proposed to effectively describe traffic flow in road networks. Nonetheless, various types of traffic data have been shown to exhibit self-similarity and burstiness. Considering these spatial and temporal correlations along with traffic dynamics bring significant potential advantages for the development of efficient traffic state prediction.

Major existing approaches, regarding the treatment of temporal and spatial features of traffic flow for short-term forecasting, first need to calibrate both temporal and spatial correlations among the available traffic data. The calibrated results are then fed to the prediction algorithm externally [27]. In addition to the extremely large amount of data and arduous calibration, this kind of method requires evaluations of both autocorrelations and cross-correlations to obtain the appropriate time lag of each correlated series [8]. These preprocessing steps, which are also computationally expensive, render the spatial and temporal features of traffic flow difficult to be incorporated in short-term traffic prediction. An alternative approach to consider the spatial and temporal correlation analysis is to utilize dynamics of the traffic model in conjunction with some extended dynamics, e.g., random walk [29], [30] and colored noise [7]. In this approach, the spatial and temporal correlations of the uncertainties are propagated by the extended dynamics (auxiliary dynamical systems), which, in turn, affects the traffic flow models by the interconnected structure.

In the SCTM case, by introducing the cross-correlations (with/without time lags) directly into the model, its Markovian property may not be retained. Moreover, most short-term traffic state prediction methods are based on 5-min (or above) aggregated traffic data. Such an interval would be too large for a dynamic traffic flow model to produce accurate estimation or prediction. On the other hand, the calibration of the autocorrelations and cross-correlations by a time increment as small as the simulation time increment of a traffic flow model will be too computationally intensive and impractical. Due to the discrete-time stochastic bilinear system formulation of the SCTM, it is convenient for us to extend the SCTM to consider the spatial and temporal correlations of demand and supply functions by the auxiliary dynamical system approach. In this paper, a multivariate normal distribution (MND)-based best linear predictor is adopted to forecast boundary variables (or demand) and supply functions by incorporating the historical temporal correlated traffic flow data and real-time detected data as inputs. Similar to the spatial correlation analysis recently developed in wireless sensor networks, we assume the samples of spatial correlation to be temporally independent or uncorrelated. This assumption reduces the complexity of the problem and ensures the separability of the spatial and temporal covariance matrices. The predicted boundary variables and supply functions are imported to the SCTM to perform short-term traffic state prediction. The independent assumption of the SCTM is relaxed by incorporating the covariance structure, which is calibrated from the spatial correlation analysis of the measured traffic flow data, for probabilistic traffic state evaluation. For real-time application purposes, prediction is conducted in a rolling horizon manner.

The rest of this paper is structured as follows. Section II reviews the SCTM and extends it to consider spatial correlations. Section III introduces a multivariate best linear predictor to extend the SCTM to consider the correlations of traffic flow for forecasting boundary variables and supply functions. Section IV introduces the concept of rolling horizon and proposes a framework for real-time prediction. Section V conducts an empirical study for traffic state prediction on a segment of I210-W. Finally, conclusions are drawn and future work is discussed in Section VI.

[1]This is a common assumption of the parametric methods. The parametric methods may produce larger prediction errors when the model parameters are biased.

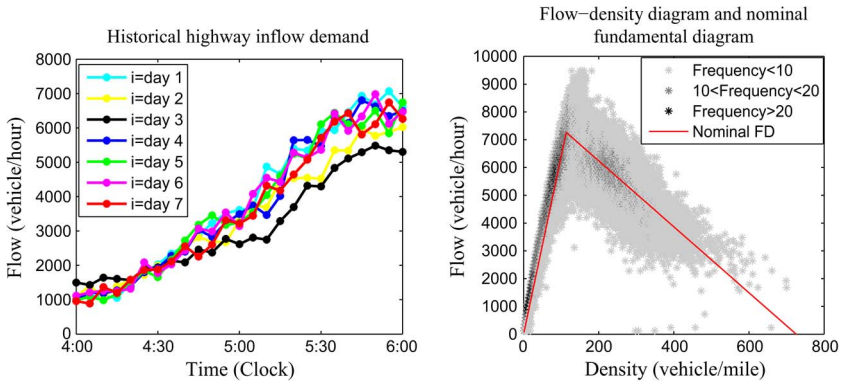

Fig. 1. Stochastic inflow and fundamental diagram.

II. Descriptions of the Stochastic Cell Transmission Model and Its Extension

The stochastic cell transmission model (SCTM) is a stochastic dynamic traffic flow model for a freeway segment proposed by Sumalee *et al.* [20] and later extended to the network case by Zhong *et al.* [32]. The SCTM considers the stochastic characteristics of the fundamental diagram (uncertain flow–density relationship) and admits the stochastic travel demand as an exogenous input, as shown in Fig. 1. For the supply side, the means and variances of different parameters (i.e., free-flow speed, jam-density, critical-density, and backward-wave speed), which govern the fundamental diagram, are calibrated based on the statistics of the observed data (e.g., the gray dots on the right-hand side in Fig. 1) for each section equipped with detectors. The SCTM defines the random inflows (uncertain demand) and random parameters of the fundamental flow–density diagram (uncertain supply functions) as boundary variables. It accepts the means and variances of the boundary variables as exogenous inputs and then calculates the means and variances of the traffic densities, outflow of the freeway segment, and probabilities of its operational modes as outputs based on the measured boundary conditions. To be specific, at each simulation time step, the stochastic flow propagation between adjacent cells under each mode can be determined by the flow–density relationship. Each of these stochastic flow profiles is associated with a certain probability corresponding to an operational mode. The flow propagated from the upstream to downstream cell in the next time step is a mixture distribution of the stochastic flows of the operational modes. The performance of the SCTM in estimating stochastic traffic densities and dynamic travel time distribution against empirical freeway traffic data was validated in [20] and [21]. Fig. 2 shows the application of the SCTM to represent a freeway segment without on-ramp or off-ramp, which is divided into two cells. Similar to the switching-mode model proposed by Muñoz *et al.* [14], five probabilistic events (we refer to them as five operational modes hereafter) are defined in the SCTM to represent the freeway state under different congestion levels, i.e., two steady-state modes, namely, the FF mode and the CC mode, and three transient modes, namely, the CF mode, the FC1 mode, and the FC2 mode. The division of the FC1 and FC2 modes depends on the moving direction of the wavefront. If the wavefront is moving downstream, then the freeway segment is

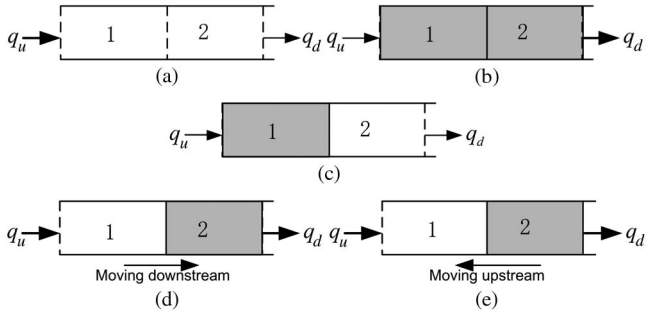

Fig. 2. Five traffic operational modes for a freeway segment with two cells.

in the FC1 mode; otherwise, it is in the FC2 mode. Due to the random demand and supply, the SCTM defines the probabilities of occurrences for these five events. Each mode of the SCTM can be represented by the following bilinear system formulation as

$$\rho(k+1) = \left(A_0 + \sum_{i=1}^{2} A_i \omega_i(k)\right)\rho(k) + \left(B_0 + \sum_{i=1}^{2} B_i \omega_i(k)\right)\lambda(k) + Bu(k) \quad (1)$$

where B, A_i, and B_i, $i = 0, 1, 2$, are constant matrices to be defined later, and $\omega_i(k), \forall k \in N$, are second-order processes consisting of mutually uncorrelated real-valued random variables.

In the FF mode, we set $\omega_i(k)$ to be the free-flow speed $v_{f,i}(k)$ in (1), and the state equation can be represented as

$$\rho(k+1) = \left(I + \sum_{i=1}^{2} A_i v_{f,i}(k)\right)\rho(k) + Bu(k) \quad (2)$$

where

$$A_1 = \begin{bmatrix} -\frac{T_s}{l_1} & 0 \\ \frac{T_s}{l_2} & 0 \end{bmatrix}$$
$$A_2 = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{T_s}{l_2} \end{bmatrix}$$
$$B = \begin{bmatrix} \frac{T_s}{l_1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Equation (2) is a special case of (1) with $B_i, i = 1, 2$ being null matrices and $\lambda(k)$ being a null vector.

In the CC mode, we define $\omega_i(k) = w_{c,i}(k)$ and the vector $\lambda(k) = (\rho_{J,1}(k), \rho_{J,2}(k))^T$. The state equation is then

$$\rho(k+1) = \left(I + \sum_{i=1}^{2} A_i w_{c,i}(k)\right)\rho(k) + \sum_{i=1}^{2} B_i w_{c,i}(k)\lambda(k) + Bu(k) \quad (3)$$

where

$$A_1 = \begin{bmatrix} -\frac{T_s}{l_1} & 0 \\ 0 & 0 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 0 & \frac{T_s}{l_1} \\ 0 & -\frac{T_s}{l_2} \end{bmatrix}$$
$$B = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{T_s}{l_2} \end{bmatrix} \qquad B_i = -A_i, \ \ i = 1, 2.$$

In the CF mode, we can define $\omega_1(k) = w_{c,1}(k)$, $\omega_2(k) = v_{f,2}(k)$, and the vector $\lambda(k) = (\rho_{J,1}(k), Q(k))^T$. The state equation is then

$$\rho(k+1) = \left(I + \sum_{i=1}^{2} A_i \omega_i(k)\right)\rho(k) + \left(B_0 + \sum_{i=1}^{2} B_i \omega_i(k)\right)\lambda(k) + Bu(k) \quad (4)$$

where

$$A_1 = \begin{bmatrix} -\frac{T_s}{l_1} & 0 \\ 0 & 0 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{T_s}{l_2} \end{bmatrix}$$
$$B_0 = \begin{bmatrix} 0 & -\frac{T_s}{l_1} \\ 0 & \frac{T_s}{l_2} \end{bmatrix}, \qquad B_1 = -A_1 \quad B_2 = 0 \quad B = 0.$$

In the FC1 mode, we define $\omega_1(k) = v_{f,1}(k)$, $\omega_2(k) = 0$, and $\lambda(k)$ as a null vector. The state equation is then

$$\rho(k+1) = (I + A_1\omega_1(k))\,\rho(k) + Bu(k) \quad (5)$$

where

$$A_1 = \begin{bmatrix} -\frac{T_s}{l_1} & 0 \\ \frac{T_s}{l_2} & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} \frac{T_s}{l_1} & 0 \\ 0 & -\frac{T_s}{l_2} \end{bmatrix}.$$

In the FC2 mode, we define $\omega_1(k) = 0$, $\omega_2(k) = w_{c,2}(k)$, and $\lambda(k) = (0, \rho_{J,2}(k))^T$. The state equation is

$$\rho(k+1) = (I + A_2\omega_2(k))\,\rho(k) + B_2\omega_2(k)\lambda(k) + Bu(k)$$

where

$$A_1 = 0, \qquad A_2 = \begin{bmatrix} 0 & \frac{T_s}{l_1} \\ 0 & -\frac{T_s}{l_2} \end{bmatrix}$$
$$B_1 = 0, \qquad B_2 = \begin{bmatrix} 0 & -\frac{T_s}{l_1} \\ 0 & \frac{T_s}{l_2} \end{bmatrix}, \qquad B = \begin{bmatrix} \frac{T_s}{l_1} & 0 \\ 0 & -\frac{T_s}{l_2} \end{bmatrix}.$$

The evaluation of the mean and variance of the given bilinear system is studied in [20]. The corresponding probabilities of occurrence of the five modes can be defined as follows:

FF mode: $P_{FF}(k) \triangleq \Pr(\rho_u(k-1) < \rho_{c,1}(k-1) \cap \rho_d(k-1) < \rho_{c,2}(k-1))$;

CC mode: $P_{CC}(k) \triangleq \Pr(\rho_u(k-1) \geq \rho_{c,1}(k-1) \cap \rho_d(k-1) \geq \rho_{c,2}(k-1))$;

CF mode: $P_{CF}(k) \triangleq \Pr(\rho_u(k-1) \geq \rho_{c,1}(k-1) \cap \rho_d(k-1) < \rho_{c,2}(k-1))$; and

FC mode: $P_{FC}(k) \triangleq 1 - (P_{FF}(k) + P_{CC}(k) + P_{CF}(k))$, with the wavefront moving downstream as

$$P_{D|FC}(k) \triangleq \Pr\left(v_{f,1}(k-1)\bar{\rho}_1(k-1) \leq w_2(k-1)\left(\rho_{J,2}(k-1) - \bar{\rho}_2(k-1)\right)\right)$$

and the wavefront moving upstream as $P_{U|FC}(k) = 1 - P_{D|FC}(k)$. Then, the probabilities of FC1 and FC2 to occur at time step k are FC1 mode: $P_{FC1}(k) \triangleq P_{D|FC}(k)P_{FC}(k)$ and FC2 mode: $P_{FC2}(k) \triangleq P_{U|FC}(k)P_{FC}(k)$, where $\rho_{c,i}$ is the critical density, w_i is the backward congestion wave speed, and $\rho_{J,i}$ is the jam density of cell i, respectively. $\bar{\rho}_i$ is the joint density of cell i, which is defined as a finite mixture distribution of the five modes. Denote the joint traffic density vector and its mean and covariance matrix as $\bar{\rho}(k)$, $E(\bar{\rho}(k)|\theta(k))$, and $\mathrm{Var}(\bar{\rho}(k)|\theta(k))$, where $\theta(k) = \{\theta_s(k)\}$, $\theta_s(k) = (\rho_s(k), P_s(k))$, and $\rho_s(k)$ denote the vector of cell densities of mode s. The probability density function (pdf) of joint traffic density $f(\bar{\rho}(k)|\theta(k))$ is defined as

$$f\left(\bar{\rho}(k)|\theta(k)\right) = \sum_s P_s(k) f\left(\bar{\rho}(k)|\theta_s(k)\right). \quad (6)$$

Under mixture model (6), expectation $E(\bar{\rho}(k)|\theta(k))$ is given by

$$E\left(\bar{\rho}(k)|\theta(k)\right) = \sum_s P_s(k) E\left(\rho_s(k)\right). \quad (7)$$

Let $\mu_s(k) = E(\rho_s(k))$ and $\mu(k) = E(\bar{\rho}(k)|\theta(k))$. Then, $\mu(k) = \sum_s P_s(k)\mu_s(k)$. To evaluate $\mathrm{Var}(\bar{\rho}(k)|\theta(k))$, we define the covariance matrix of $\rho_s(k)$ as

$$\psi_s(k) = E\left(\left(\rho_s(k) - \mu_s(k)\right)\left(\rho_s(k) - \mu_s(k)\right)^T\right).$$

Then, covariance matrix $\mathrm{Var}(\bar{\rho}(k)|\theta(k))$ can be evaluated as

$$\mathrm{Var}\left(\bar{\rho}(k)|\theta(k)\right) = \sum_s P_s(k)\left(\psi_s(k) + \mu_s(k)\mu_s^T(k)\right) - \mu(k)\mu^T(k). \quad (8)$$

In the original SCTM, the uncorrelated assumption is enforced to simplify the probability evaluation and traffic flow propagation. This uncorrelated assumption is no longer valid when the spatial–temporal correlations are incorporated into the model, which, in turn, affects two major components of the SCTM, i.e., evaluations of probabilities of occurrence of different modes and traffic flow propagation. Here, we investigate the scenarios introduced in [20, Sec. 3.3]. To begin with, we define X_{ud} as $X_{ud} = [X_u, X_d]^T = [\tilde{\rho}_u(k) - \rho_{c,1}(k), \tilde{\rho}_d(k) - \rho_{c,2}(k)]^T$, where we have omitted time index k to X_{ud} to save notation. Then, the pdf of X_{ud} is a bivariate normal distribution

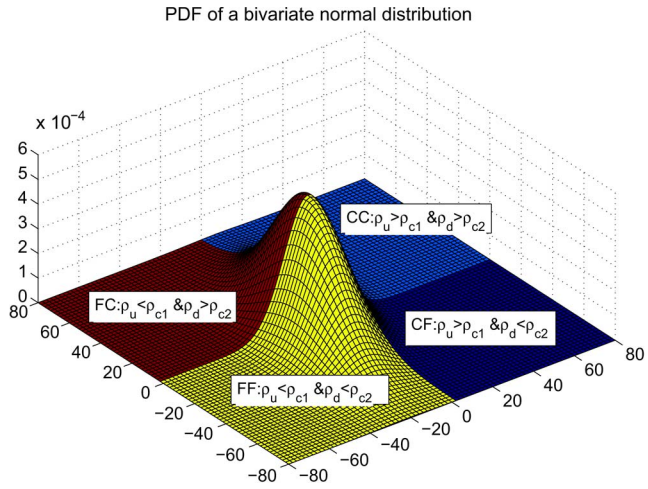

Fig. 3. PDF and probabilities of occurrence of modes.

as all its components are assumed to be normally distributed, i.e.,

$$\text{pdf}(X_{ud};\mu_{ud},\Sigma_{ud}) = \frac{1}{(2\pi)|\Sigma_{ud}|^{1/2}} e^{-Q(X_{ud};\mu_{ud},\Sigma_{ud})/2} \quad (9)$$

where $\mu_{ud} = (\mu_u,\mu_d)^T = (\bar{\rho}_u - \bar{\rho}_{c,1}, \bar{\rho}_d - \bar{\rho}_{c,2})^T$ is the expectation of X_{ud}, and Σ_{ud} is the covariance matrix, with

$$\Sigma_{ud} = \begin{bmatrix} \text{Var}(X_u) & \text{Cov}(X_u, X_d) \\ \text{Cov}(X_u, X_d) & \text{Var}(X_d) \end{bmatrix}$$

and $Q(X_{ud};\mu_{ud},\Sigma_{ud}) = (X_{ud}-\mu_{ud})^T \sum_{ud}^{-1}(X_{ud}-\mu_{ud})$. The probability of occurrence of the FF mode is now

$$\begin{aligned} P_{FF}(k) &\triangleq \Pr\left(\tilde{\rho}_u(k) < \rho_{c,1} \cap \tilde{\rho}_d(k) < \rho_{c,2}\right) \\ &= \Pr(X_u < 0 \cap X_d < 0) \\ &= \int_{-\infty}^{0}\int_{-\infty}^{0} \text{pdf}(X_{ud};\mu_{ud},\Sigma_{ud}) dX_u dX_d. \end{aligned} \quad (10)$$

The probabilities of occurrence of the other modes can be similarly defined as

$$P_{CC}(k) = \int_{0}^{+\infty}\int_{0}^{+\infty} \text{pdf}(X_{ud};\mu_{ud},\Sigma_{ud}) dX_u dX_d \quad (11)$$

$$P_{CF}(k) = \int_{0}^{+\infty}\int_{-\infty}^{0} \text{pdf}(X_{ud};\mu_{ud},\Sigma_{ud}) dX_u dX_d \quad (12)$$

$$P_{FC}(k) = 1 - \left(P_{FF}(k) + P_{CC}(k) + P_{CF}(k)\right). \quad (13)$$

A demonstration of the pdf of the bivariate normal distribution previously discussed is given in Fig. 3. The probabilities of occurrence of different modes can be evaluated by the area of the corresponding regimes, e.g., the probability of the FF mode is given by the yellow cover. Evaluation of the autocorrelation matrix of the bilinear system should be adjusted accordingly. For example, the covariance of two supply functions is [20, eq. (29)], i.e., $\text{Cov}(\omega_{s,i}(k),\omega_{s,j}(k)) = E(\omega_{s,i}(k)\omega_{s,j}(k)) - E(\omega_{s,i}(k))E(\omega_{s,j}(k)), i \neq j$, which is now nontrivial due to the spatially correlated assumption of the supply functions. For the purpose of simulation of traffic dynamics, it can be achieved by solving the recursive equations using a computer program. Therefore, we are not going to dwell upon all of the analytical equations in detail due to limited space. Interested readers can refer to [35] for detailed derivation of mean and variance equations of traffic dynamics incorporating the spatial correlations of supply functions with application to optimal and robust strategies for freeway traffic management under demand and supply uncertainties.

III. Best Linear Predictor and Forecasting of Flow Profiles and Supply Functions

The accuracy of prediction heavily depends on the similarity between the historical data and the current trend of the traffic flow when we apply the original SCTM to predict traffic state. Prediction may fail when the traffic network is suffering from abnormal traffic conditions (see, e.g., the empirical study in Section V). Prediction accuracy would increase by incorporating the spatial and temporal correlations of traffic flow into the SCTM. To extend the SCTM for traffic state prediction, we may adopt a certain prediction algorithm to forecast the inflow/outflow profiles and supply functions. Here, the MND-based best linear predictor [26] is utilized to accomplish this objective.

Take the prediction of a flow profile at a given location as an example, we assume that the dynamic statistical flow profiles are normally distributed. The flow profile vector conditioned on the current measurement can be forecasted by the following best linear predictor:

$$E(q_f|q_m) = E(\hat{q}_f) + s_{\hat{q}_{f,m}} \Sigma_{\hat{q}_m}^{-1} \left(q_m - E(\hat{q}_m)\right) \quad (14)$$

where $q_f = \text{col}(q(k+1),\ldots,q(k+n))$ denotes the extended flow vector to be forecasted for the coming predicting horizon, and $\hat{q}_f$ is the corresponding historical data during the same period, with $E(\hat{q}_f)$ denoting the mean value of $\hat{q}_f$, whereas $q_m = \text{col}(q((k-l+1)),\ldots,q(k))$ is the measured flow vector for the current time period. $s_{\hat{q}_{f,m}} = (s_{i,j})_{n\times l}$ denotes the covariance matrix between q_f and q_m with

$$s_{i,j} = \text{Cov}\left(\hat{q}_f(i),\hat{q}_m(j)\right) = \text{Cov}\left(\hat{q}(k+i),\hat{q}(k-l+j)\right)$$

$i = 1,\ldots,n;\ j = 1,\ldots,l$, and $\Sigma_{\hat{q}_m}^{-1}$ is the generalized inverse of the covariance matrix of $\hat{q}_m$ with $\Sigma_{\hat{q}_m} = (\sigma_{i,j})_{l\times l}$, and

$$\sigma_{i,j} = \text{Cov}\left(\hat{q}_m(i),\hat{q}_m(j)\right) = \text{Cov}\left(\hat{q}(k-l+i),\hat{q}(k-l+j)\right)$$

$i, j = 1,\ldots,l$. The covariance matrix of the flow vector to be predicted can be evaluated as

$$\text{Cov}(q_f|q_m) = \text{Cov}(\hat{q}_f) - s_{\hat{q}_{f,m}} \Sigma_{\hat{q}_m}^{-1} s_{\hat{q}_{f,m}}^T. \quad (15)$$

The prediction of supply functions conditioned on the current observation can be also predicted by the same algorithm, i.e.,

$$E(\delta_f|\delta_m) = E(\hat{\delta}_f) + s_{\hat{\delta}_{f,m}} \Sigma_{\hat{\delta}_m}^{-1} \left(\delta_m - E(\hat{\delta}_m)\right) \quad (16)$$

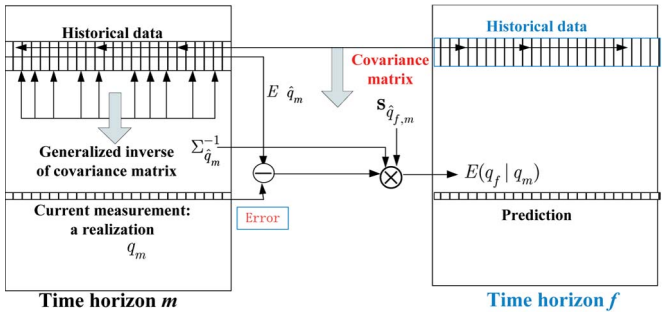

Fig. 4. Mechanism of the best linear predictor.

where $E(\delta_f|\delta_m)$ is the mean value of parameter δ_f given the current detection of δ_m, and $s_{\hat{\delta}_{f,m}}$ and $\Sigma_{\hat{\delta}_m}$ are defined as those in (14). Similarly, the prediction of the covariance matrix can be defined as

$$\mathrm{Cov}(\delta_f|\delta_m) = \mathrm{Cov}(\hat{\delta}_f) - s_{\hat{\delta}_{f,m}} \Sigma^{-1}_{\hat{\delta}_m} s^T_{\hat{\delta}_{f,m}}. \tag{17}$$

The mechanism of the best linear predictor is shown in Fig. 4. We utilize the difference between the historical mean and current measurement of the flow profile (we will refer to this as "error") to correct the prediction. Ideally, if this error is zero, the predicted flow profile "is equal to" the historical flow profile, or no adjustment is made. Otherwise, we adjust the prediction by the error weighted by the covariance matrices (or their inverses) if the error is not zero. The best linear predictor can forecast the flow profiles and supply functions by considering their historical statistics and the real-time detection. This method avoids time-consuming approaches such as data mining in database (e.g., k-NN algorithm). Thus, it is efficient. Adopting time-dependent supply functions instead of the fixed parameters renders the prediction of traffic state more reliable.

Remark 3.1: The linear predictor adopted here is based on the experience of California Partners for Advanced Transportation Technology (see, e.g., [6]). The reason is that occupancies and volumes obtained from detectors at nearby locations are highly correlated. Therefore, measurements from one location can be used to estimate quantities at other locations, and a more accurate estimate can be formed if all the neighboring detectors are used in the estimation. As deemed by Chen *et al.* in [6], "The high correlation among neighboring measurements means that linear regression is a good way to predict one from the other. It is also easy to implement and fast to run."

IV. Framework of Online Traffic State Prediction

Recognizing the fact that measurement systems always suffer from inherent bad detection, errors, and communication delays, it may not be possible for us to implement a traffic simulator in real time. This may be overcome by using this rolling horizon approach, as shown in Fig. 5. The measure data are not directly used by the simulator, e.g., the SCTM in this paper. On the other hand, it is processed by a predictor/filter, e.g., data filtering (imputation of bad detections) and the prediction of boundary variables and supply functions, as discussed in the previous sections. The filtered data and predicted boundary variables and supply functions are stored in a repository of road networks, which is a database. The simulator fetches data directly from this repository to conduct real-time traffic state estimation prediction.

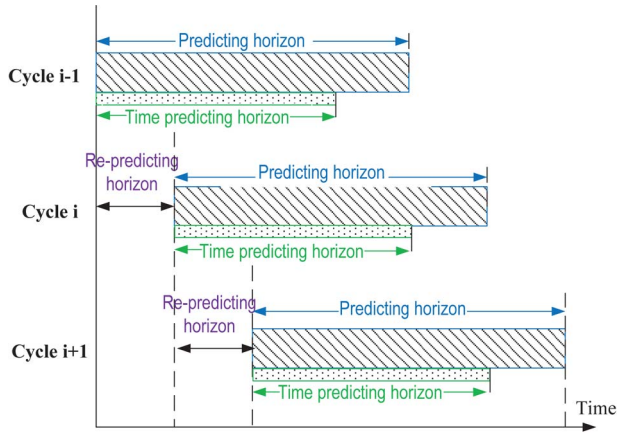

Fig. 5. Framework of the SCTM-based traffic prediction with rolling horizon.

The forecasting framework utilizes the rolling horizon approach, which is a concept that is widely used in online predictions. In this rolling horizon approach, traffic density and journey time are predicted cycle by cycle with predefined prediction horizon.[2] Every two neighboring horizons are differentiated by repredicting horizon, which is generally shorter than predicting horizon, as shown in Fig. 5. The rolling horizon approach can be interpreted as, for example, the route guidance in an ATIS, under this model, the route guidance is periodically generated and evaluated for a given time horizon (or the prediction horizon, e.g., 45 min) based on the latest information available. However, 45 min may be too long for us to trust the prediction (or the prediction may not reflect the current traffic state). On the other hand, the measurement systems keep detecting the real-time traffic state, e.g., the Performance Measurement System (PeMS) can provide real-time traffic information every 30 s and aggregate traffic state every 5 min. We would like to use this information, e.g., the 5-min aggregate data, to adjust the prediction, which has been discussed in the previous sections. Therefore, the current guidance until the next guidance update (e.g., 5 min) is actually implemented. Travelers may update their route choices according to the guidance updates. We call this interval of 5 min as the rolling horizon step size.[3] A detailed flowchart implementation of the proposed traffic state prediction framework is shown in Fig. 6.

V. Empirical Study: Traffic State Prediction on a Short Segment of I210-W

In this empirical study, we will validate the proposed traffic state prediction framework on the test site adopted in [20], i.e.,

[2]This is also known as rolling horizon length, which is the time period for which prediction takes place. This length is a function of the maximum trip length. A rolling horizon is usually divided into short time intervals, e.g., rolling horizon step size that specifies how often guidance is renewed.

[3]This step size may be determined based on the level of variation in traffic conditions over time (e.g., a longer step size can be used if traffic conditions do not change very much), the resolution of the measurement system, timing of incidents, and available computational resources.

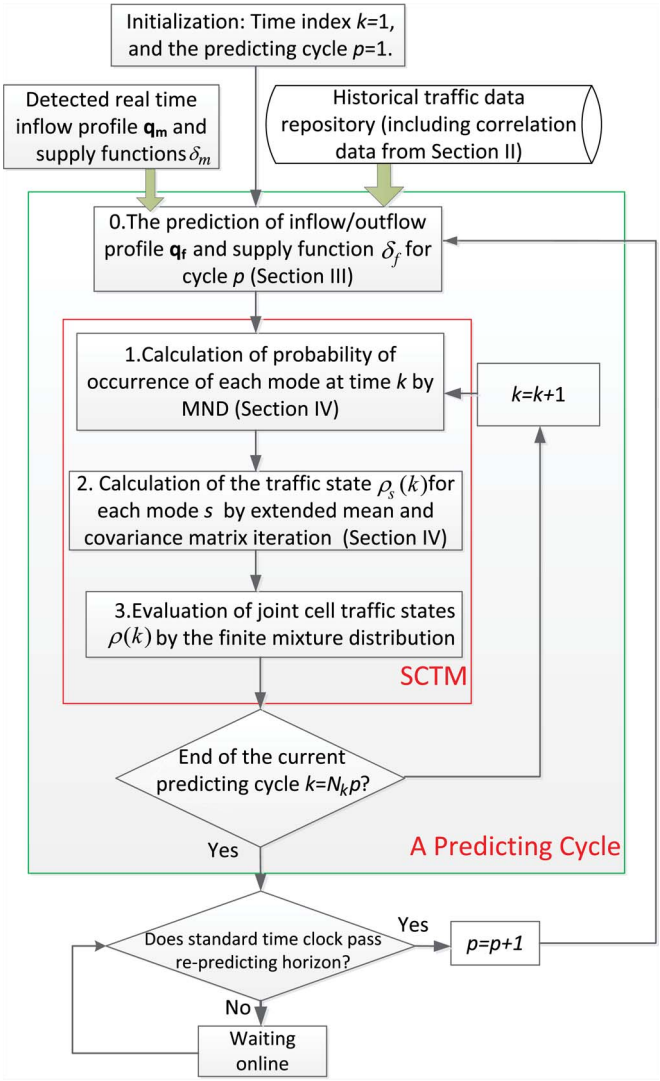

Fig. 6. Flowchart of traffic state prediction.

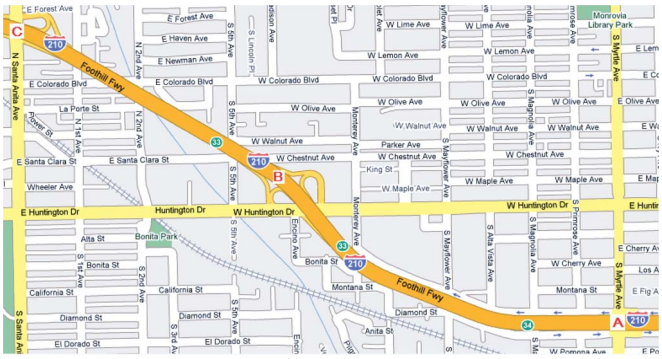

Fig. 7. Map of the test site (source: Google map).

a 2-mi segment of the Interstate 210 Freeway near Los Angeles, CA, USA.

A. Test Site Description and Model Parameter Calibration

A segment of the Interstate 210 West bound, which is approximately 2 mi in length, is chosen in this case study, as shown in Fig. 7. This short segment, which is located in Monrovia, CA, USA, stretches from S Myrtle Ave (A) through W. Huntington Dr (B) to N Santa Anita Ave (C) with two on-ramps and two off-ramps. This segment of the freeway is chosen here for the following reasons.

1) The high level of recurrent congestion within the section can be observed in the early morning period (6 A.M.–10 A.M.).
2) The segment possesses necessary infrastructure and traffic detectors embedded in the on-ramps and mainline lanes for data collection.

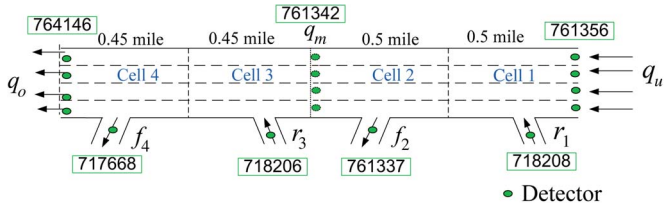

Fig. 8. Section of I210-W divided into four cells and its detector configuration.

TABLE I
SIMULATION SETTINGS

Items	Time interval
Raw detection time interval of PeMS	30 seconds
Aggregated time interval of PeMS	5 minutes
Simulation time interval	5 seconds
Predicting horizon	30 minutes
Rolling horizon step size	5 minutes

This section is instrumented with single-loop inductance detectors, which are embedded in the pavement along the mainline, high-occupancy vehicle lane, on-ramps, and off-ramps. Each loop detector provides raw data such as traffic volume (veh/time-step) and occupancy (%) for the corresponding lane for 30 s, where, in this empirical study, the first item is conversed to general (for all lanes) traffic flow rate (veh/h), and the second item becomes general densities (veh/mi) via dividing the occupancy by g-factor (ft/veh), which is the effective vehicle length. For the convenience of analyzing the prediction of boundary variables and supply functions, we present a schematic representation of the freeway segment and its detector configuration in Fig. 8. Boundary variables of this freeway segment, i.e., q_u, r_1, f_1, q_o, r_2, and f_2, are predicted using the best linear predictor. Here, we pretend that the detector in the middle, i.e., q_m, is missing (i.e., we do not use it in the simulation) for validation of the performance of the proposed method.

Traffic flow data of 7 h (4:00 A.M.–11:00 A.M.) collected on Tuesday, Wednesday, and Thursday in March and April of 2008 and 2009 provided by the PeMS are used in this test to calculate spatial and temporal correlations of the related traffic characteristics. Traffic data collected on the morning of March 3, 2011 (normal scenario) and March 8, 2011 (abnormal scenario) are selected to test the performance of the proposed framework under different traffic scenarios. In this empirical study, the spatial–temporal correlations are calibrated using the methods depicted in Section V-B. Basic configurations of the simulation are outlined in Table I. Predicting the horizon length is assigned to be 30 min. Rolling horizon step size,

i.e., repredicting horizon, is 5 min, which can be adjusted according to the accuracy requirement and real-time measured data resolution.

B. Spatial–Temporal Correlation Phenomena: Data Preparation

Spatial–temporal correlations can be utilized to predict short-term traffic state. Here, we briefly introduce the mathematical definitions of the correlations utilized in this empirical study. Temporal covariance and correlation coefficient of $q(k)$ between two adjacent time steps k and $k-z$ are defined as

$$\mathrm{Cov}_{tm}\left(q(k), q(k-z)\right) = \frac{\sum_{i=1}^{N}\left(q_i(k)-\bar{q}(k)\right)\left(q_i(k-z)-\bar{q}(k-z)\right)}{N-1} \tag{18}$$

$$\gamma_{tm}\left(q(k), q(k-z)\right) = \frac{\mathrm{Cov}_{tm}\left(q(k), q(k-z)\right)}{s_{q(k)}s_{q(k-z)}}, \quad z=1,2,\ldots,l_e,\ l_e \le k-1 \tag{19}$$

where $q_i(k)$ is the value of the detected flow during interval k on the ith day; z is a positive integer that is less than or equal to l_s, which is a predefined bound (also an integer) for the calculation; and $s_{q(k)}$ denotes the standard deviation of detected flow $q(k)$ for N sample days.

Temporal correlation of the parameters of a fundamental diagram (or supply functions) for a given location can be also similarly analyzed, i.e.,

$$\mathrm{Cov}_{tm}\left(\delta(k_p), \delta(k_p-l_p)\right) = \frac{\sum_{i=1}^{N}\left(\delta_i(k_p)-\bar{\delta}(k_p)\right)\left(\delta_i(k_p-l_p)-\bar{\delta}(k_p-l_p)\right)}{N-1} \tag{20}$$

$$\gamma_{tm}\left(\delta(k_p), \delta(k_p-l_p)\right) = \frac{\mathrm{Cov}_{tm}\left(\delta(k_p), \delta(k_p-l_p)\right)}{s_{\delta(k_p)}s_{\delta(k_p-l_p)}},\ l_p=1,2,\ldots,l_{p,\varepsilon},\ l_{p,\varepsilon}<k_p \tag{21}$$

where $\delta_i(k_p)$ is a certain parameter calibrated with data detected during time period k_p (e.g., half an hour) on the ith day. $\bar{\delta}(k_p)$ is the mean of $\delta(k_p)$ and $s_\delta(k_p)$ is the corresponding standard deviation over N days.

The spatial covariance and correlation coefficient of adjacent supply functions during time interval k_p are thus defined as

$$\mathrm{Cov}_{\mathrm{sp}}\left(\delta_m(k_p), \delta_n(k_p)\right) = \frac{\sum_{i=1}^{N}\left(\delta_{m,i}(k_p)-\bar{\delta}_m(k_p)\right)\left(\delta_{n,i}(k_p)-\bar{\delta}_n(k_p)\right)}{N-1} \tag{22}$$

$$\gamma_{\mathrm{sp}}\left(\delta_m(k_p), \delta_n(k_p)\right) = \frac{\mathrm{Cov}_{\mathrm{sp}}\left(\delta_m(k_p), \delta_n(k_p)\right)}{s_{\delta_m(k_p)}s_{\delta_n(k_p)}} \tag{23}$$

where the notations in (22) and (23) are similar with the notations in (20) and (21) but for the differentiation of $\delta_m(k_p)$ and $\delta_n(k_p)$ denoted by m and n, which represent the adjacent locations. Spatial correlation of densities can be evaluated as

$$\mathrm{Cov}_{\mathrm{sp}}\left(\rho_m(k), \rho_n(k)\right) = \frac{\sum_{i=1}^{N}\left(\rho_{m,i}(k)-\bar{\rho}_m(k)\right)\left(\rho_{n,i}(k)-\bar{\rho}_n(k)\right)}{N-1} \tag{24}$$

$$\gamma_{\mathrm{sp}}\left(\rho_m(k), \rho_n(k)\right) = \frac{\mathrm{Cov}_{\mathrm{sp}}\left(\rho_m(k), \rho_n(k)\right)}{s_{\rho_m(k)}s_{\rho_n(k)}} \tag{25}$$

where $\rho_{m,i}(k)$ is the traffic density detected at location m during time interval k on the ith day, $\bar{\rho}_m(k)$ is the mean density over N days, and $s_{\rho_m(k)}$ is the corresponding standard deviation.[4]

C. Traffic State Prediction Under Normal Traffic Condition

We first test the proposed algorithm for traffic state prediction under normal traffic condition. By normal traffic condition, we mean the traffic pattern under consideration is similar to the statistics of the historical data, as shown in Fig. 9. Note that due to the rolling horizon architecture, only the first 5 min traffic state of each prediction horizon is shown in the figure as others are overlapped by their coming prediction horizons.

As shown in Fig. 9, the predicted inflow function to the first cell is close to the measured one at the first beginning, i.e., the morning free-flow period. When the traffic condition turned to congested, demand from the upstream may not be received by the first cell. Therefore, we need to predict the inflow profile by considering the receiving function of the first cell. As demonstrated in the figure, the mean value of the prediction of the inflow profile is close to the measurement (except for the "unexpected" congestion period caused by a moving bottleneck, which will be discussed in detail later) that reflects satisfactory performance of the best linear predictor. Similarly, the outflow profile from the downstream boundary of the segment is predicted, which is not shown in the figure due to the space limit.

These flow profiles are loaded into the SCTM for traffic state prediction. Results are shown in Fig. 9 against the measured densities of the day and historical mean of all the sample days (which is not utilized in the simulation). As reflected by the figure, the prediction captures the trend of traffic dynamics rather than the historical mean. The overall mean absolute percentage error (MAPE) of all prediction horizons is around 14%, whereas the MAPE for all the repredicting horizons (i.e., the first 5 min of each predicting horizon, as shown in Fig. 9) is about 10.8% compared with the measurement from the middle detector. Most of the measured traffic data (except some points caused by the moving bottleneck) lie in the 99.7% confidence interval (i.e., three standard deviations from the mean). This

[4]More details of the analysis on spatial–temporal correlations is in the online electronic companion to this paper, which is available at http://sites.google.com/site/luciatlpan767/home/predictiontmspcrieee.

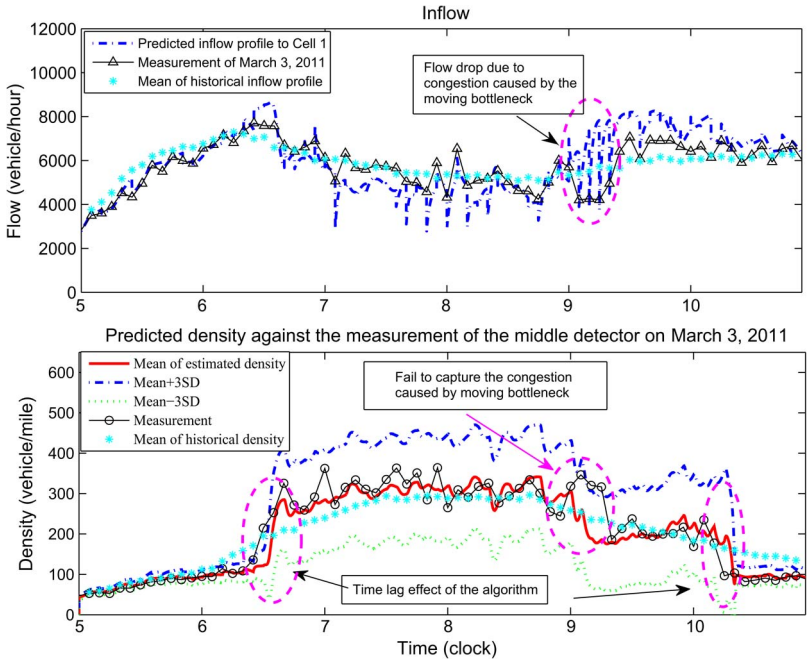

Fig. 9. Predicted flow and density against the measurements.

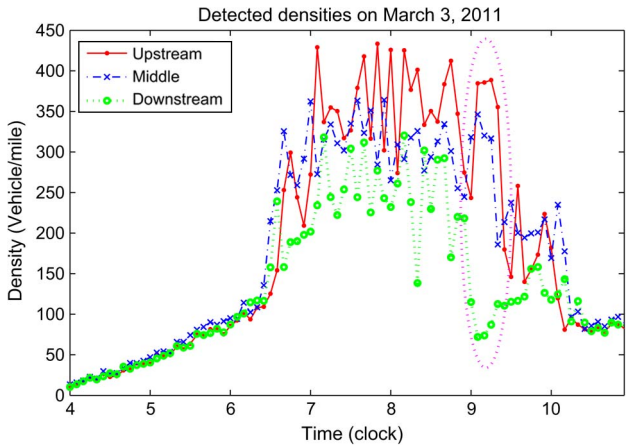

Fig. 10. Comparison of the detected densities.

example proves that the linear best predictor in conjunction with the SCTM by considering the spatial–temporal correlations is a good way for prediction in the case, which is also due to the fact that there is high correlation among neighboring measurements, as indicated in Remark 3.1. The proposed method is also easy to implement and fast to run. Nevertheless, we would like to point out some drawbacks of the proposed framework. Note that there is congestion near 9:00 A.M. that is not captured by the algorithm. Comparing the measured cell densities shown in Fig. 10, we find that this congestion was formed within the freeway segment and spilled back to the upstream cells since the detector at the downstream did not observe this congestion while the detectors at the upstream and the middle did observe it and with a time lag. Moreover, the detector at the downstream admitted an abrupt decrease in both flow and density around 9:00 A.M. and returned to the normal level in a short time. In the meantime, the congestion observed by the other detectors was dissolved. Therefore, we concluded that this congestion was caused by the moving bottleneck.

As we do not use the data from the middle detector while predicting the inflow and outflow profiles and because the prediction of inflow/outflow profiles cannot be captured (due to the fact that the observations are similar to the historical data) as well as the fact that the supply functions do not significantly change, the proposed algorithm cannot capture this congestion, as indicated in Fig. 9. After the change in the upstream inflow profile is identified by the predictor, modification of the predicted flow is made accordingly. However, this correction is too late to benefit the traffic state prediction by the SCTM, as shown in Fig. 9. This can be regarded as a drawback of this kind of reactive traffic state prediction framework (which has been reported in transportation literature), i.e., it requires sufficient information to identify the change in traffic flow patterns and to react to the change identified, which renders a certain time lag. However, the algorithm fails to capture this congestion also due to the fact that the CTM/SCTM framework cannot capture the effects of the moving bottleneck. We have tried to tackle these drawbacks of the proposed algorithm in two of our recent works, i.e., [33] (on the reactive part) and [34] (on the moving bottleneck part).

D. Traffic State Prediction Under Abnormal Traffic Condition

This section conducts an empirical study to validate the prediction results under abnormal traffic conditions by introducing spatial and temporal correlations to the short-term traffic state forecasting framework, as summarized in Fig. 5. On the morning of March 8, 2011, there was an abnormal traffic condition due to a certain incident, as shown in Fig. 11(a) and (b). As we can observe, the detected inflow/outflow profiles and the density of the day, compared with the distribution of flows and density collected over 54 sample days, admitted sudden declines (increase in the density) at around 6:00 A.M. Note also that the outflow of the segment admitted a decline

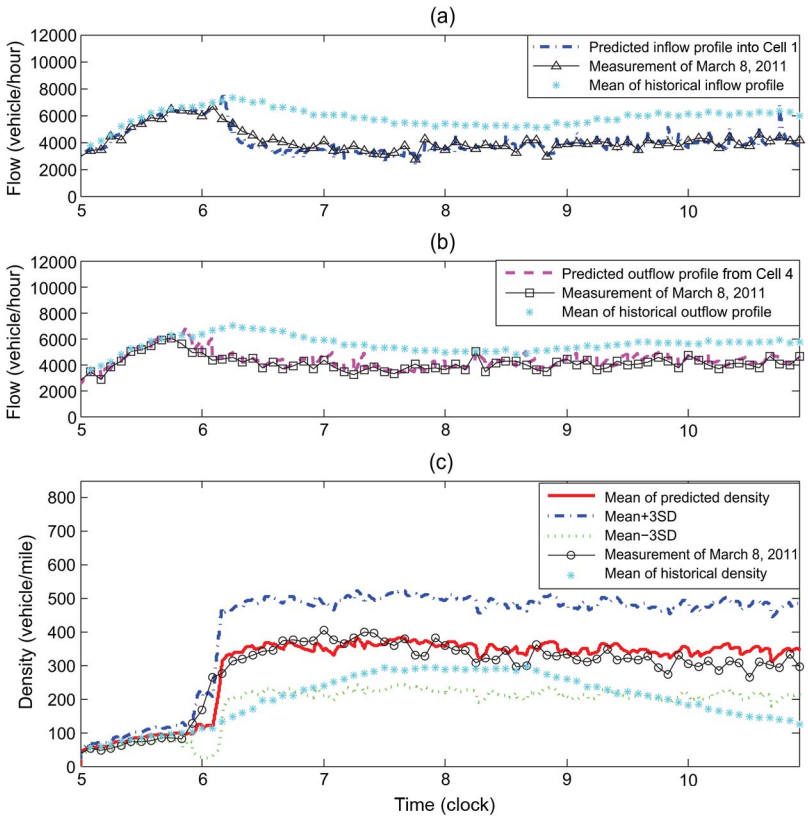

Fig. 11. Predicted flow and density against the measurements.

before the inflow did, which implies that this decline was due to congestion spill-back. Clearly, the detected traffic flows and densities thereafter were so different from the mean flows and detected densities of the 54 sample days. If we just directly input the statistical flow profiles of the 54 sample days to a traffic simulator for short-term prediction, the results may be poor in terms of accuracy due to the great difference of the flow and density profiles. Fig. 11(a) and (b) indicates that the prediction of inflow/outflow profiles is accurate, except for the time steps near the abrupt change. To see more details of the prediction mechanism, we depict the predicted inflow profiles for three rolling horizon steps near the abrupt changes in the flow profile. As shown in Fig. 12, in the first rolling horizon step, the inflow prediction is far from the actual detections (Cycle A). This is because the abnormal traffic condition occurs after the prediction started at 5:55 A.M. The prediction cannot make use of the detected flow pattern until the next rolling horizon step (when the real-time measurement is adopted to adjust the prediction), i.e., 5 min later. Huge errors between the historical flow patterns and the real-time measurements are identified by the predictor. The predictor then adjusts the prediction when the next rolling horizon step begins. However, the decrease in predicted inflow is still unsatisfactory throughout Cycle B. The prediction will keep on adjusting if the error between the historical mean and the current measurement is not zero. Cycle C (6:05 A.M.–6:35 A.M.) demonstrates this adjustment by the best linear predictor by considering temporal correlations. The inflow profile prediction becomes close to the detected one after two adjustments.

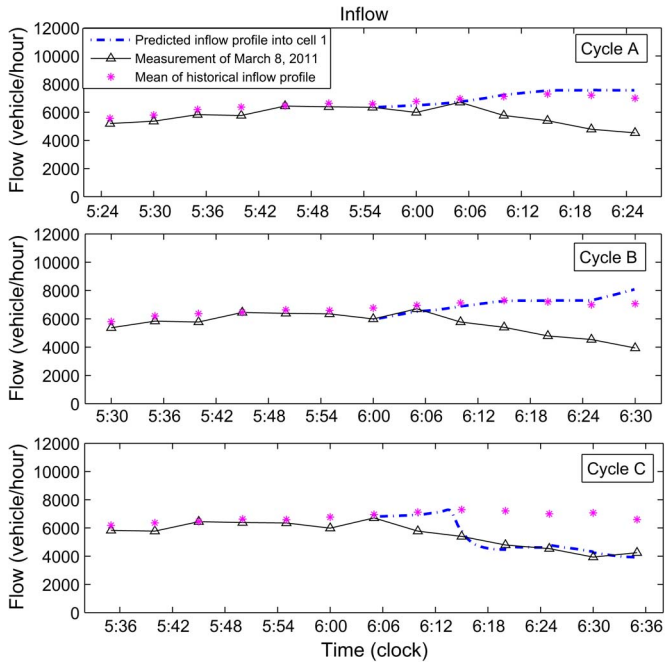

Fig. 12. Predicted inflow profile adjustment process.

It is of course that the level of improvement depends on the relationship between the characteristics of traffic flow and the rolling horizon step size (which depends on the resolution of real-time measurement and the preference of the traffic manager). One may use smaller rolling horizon step size (repredicting horizon) to accelerate the rate at which the prediction adapts

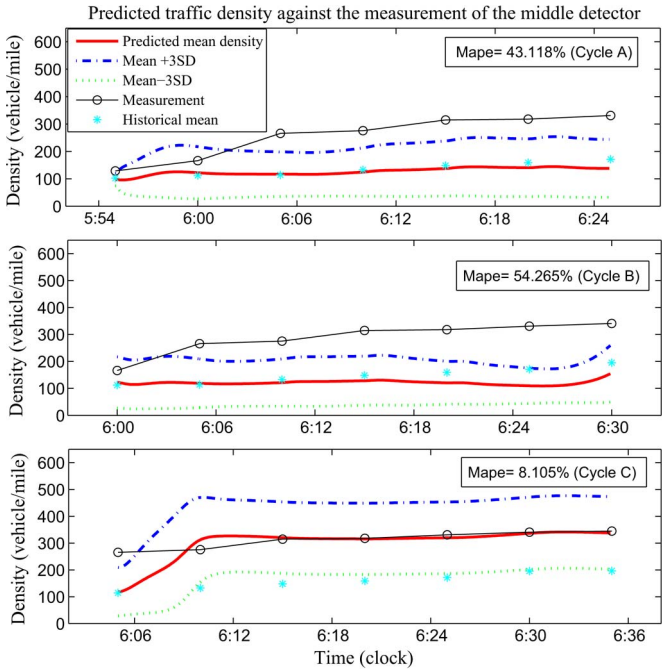

Fig. 13. Predicted traffic density adjustment process.

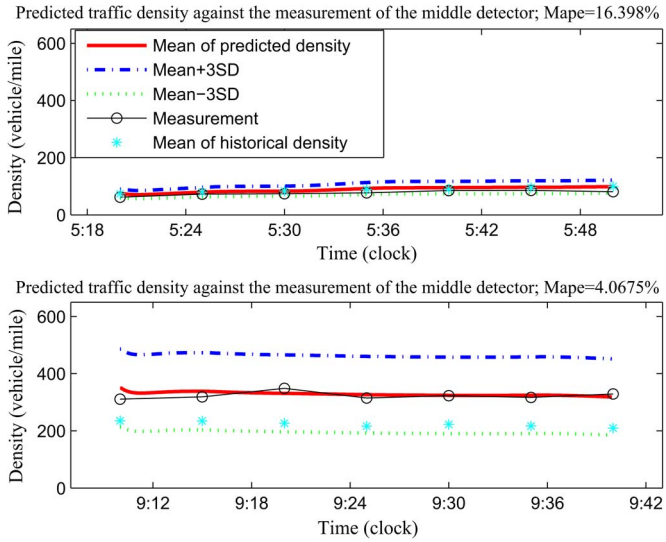

Fig. 14. Samples of prediction horizons.

to the real-time feed in traffic conditions, e.g., from 5 to 1 min (with a minimum frequency determined by the resolution of the measurement system, e.g., 30 s in the PeMS). However, this is perhaps not practical due to the sophisticated data processing, computation effort, and deployment constraints in ATIS and ATMS applications. Forecasted traffic densities of these three prediction horizons are depicted accordingly. The fundamental diagram has no significant change due to a normal weather condition and the fact that the incident does not occur on the underlying segment (as we have mentioned that the change in the flow pattern is due to congestion spill-back).

Predicted traffic density is plotted against the measurement from the middle detector for these three prediction horizons in Fig. 13. In Cycle A, the prediction underestimates traffic densities significantly. This is due to the fact that the abnormal scenario is initiated at this step, the prediction is mainly provided by the statistics of historical data. The forecasting algorithm need to wait until the real-time measurement is loaded into the predictor in the coming step to realize the error between the prediction and the detection, which then makes corrections on flow and/or supply functions forecasting. Based on the corrected flow and/or supply functions from the best linear predictor, the forecasting traffic density is adjusted by the SCTM, as demonstrated in Fig. 13. However, there is still significant error, which indicates that the adjustment is not successful. Therefore, the algorithm keeps on adjusting the predicted inflow/outflow flows (or boundary conditions), and the SCTM keeps on tuning the predicted traffic states. In Cycle C, the prediction almost coincides with the measurement, as shown in Fig. 13. Samples of prediction results for the horizons with stationary traffic conditions, e.g., free flowing and congested conditions, are shown in Fig. 14, which validate that good performance can be achieved by the proposed traffic prediction framework regardless of the great difference between historical data and measured data. This is because the tuning of the predictor with respect to the difference between the historical data and the measurement also tends to stationary some time after the traffic state turned to stationary.

Fig. 11(c) shows the effective traffic density prediction of all rolling horizon steps with the same legends as those in Fig. 9. A time lag effect can be observed in Fig. 11(c), which is due to the reactive nature of the algorithm. The time lag can be smaller by choosing smaller repredicting horizon subject to the resolution of measured traffic data, as previously discussed. For this special weekday with a long-lasting traffic incident during the morning rush hour, the prediction can capture the trend of traffic state evolution in a satisfactory manner. The overall MAPE of all prediction horizons is around 16.2%, whereas the MAPE for all repredicting horizons (i.e., the first 5 min of each predicting horizon, as shown in Fig. 9) is about 14.8% compared with the measurement from the middle detector.

Finally, we would like to point out that the choice of l in (14) is determined by balancing the accuracy of prediction and computational time. We do not have sound theoretical guidance on how to choose reasonable l. Pan [16] provided some practical investigation on this issue by empirical studies based on the criterion of MAPE variation with different values of l on the condition of $n = 6$, i.e., 30 min predicting horizon. Generally speaking, a proper choice of the length of measurement for prediction should be close to the prediction horizon, i.e., $l \approx n$. That is to say, a reasonable length of observation of real-time trend of traffic flow will benefit the prediction. Choosing a small value may not provide sufficient information for the correction of prediction. By contrast, choosing vary large l, i.e., placing too much consideration on the period far away from current time, may decrease the sensitivity of prediction. Additionally, computational time increases as l increases.

VI. Conclusion and Future Works

This paper has developed an online framework by extending the SCTM to consider the spatial–temporal correlations of uncertain flow profiles and supply functions for short-term traffic state prediction. Traffic inflow/outflow profiles and supply

functions, conditioned on the current observation, were predicted via the MND-based best linear predictor given the statistical correlations of boundary variables and supply functions, as analyzed from the historical traffic data. The predictor was able to adjust the prediction by capturing the error between the real-time measurements and the historical mean. The predicted boundary variables and supply were taken as inputs to the SCTM to perform short-term traffic state prediction. Meanwhile, the key processes of the SCTM, i.e., the evaluation of probabilities and propagation of traffic states, were improved by considering the spatial correlations of supply and traffic cell densities. For real-time or online application, the prediction was conducted in a rolling horizon manner. The extended SCTM model and the prediction algorithm were tested with empirical data from the PeMS database.

As validated by empirical studies, the proposed method performs well under different scenarios. Empirical results also revealed the potential application of the proposed method for traffic state prediction under abnormal traffic conditions, e.g., incidents and adverse weather conditions. However, the proposed algorithm suffers from the time lag effect due to its reactive nature and cannot capture the moving-bottleneck effect due to the limitation of the CTM/SCTM framework. Future efforts should be dedicated to overcoming these drawbacks. Our next step is to tackle these challenges by extending two of our recent works, i.e., [33] and [34], which intend to identify anomalies and the moving-bottleneck effect caused by lane-changing maneuvers. Another interesting aspect is to incorporate traffic state prediction into the design of freeway traffic management strategies to support real-time applications as outlined in [35].

ACKNOWLEDGMENT

The authors would like to thank California Partners for Advanced Transportation Technology for providing the traffic data from the PeMS.

REFERENCES

[1] C. Chan, T. Dillon, J. Singh, and E. Chang, "Neural-network-based models for short-term traffic flow forecasting using a hybrid exponential smoothing and Levenberg–Marquardt algorithm," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 644–654, Jun. 2012.

[2] K. Chan, S. Khadem, T. Dillon, V. Palade, J. Singh, and E. Chang, "Selection of significant on-road sensor data for short-term traffic flow forecasting using the Taguchi method," *IEEE Trans. Ind. Informat.*, vol. 8, no. 2, pp. 255–266, May 2012.

[3] K. Chan, T. Dillon, and E. Chang, "A intelligent particle swarm optimization for short-term traffic flow forecasting using on-road sensor systems," *IEEE Trans. Ind. Electron.*, vol. 60, no. 10, pp. 4714–4725, Oct. 2013.

[4] K. Chan, T. Dillon, V. Palade, E. Chang, and J. Singh, "Prediction of short-term traffic variables using intelligent swarm-based neural networks," *IEEE Trans. Control Syst. Technol.*, vol. 21, no. 1, pp. 263–274, Jan. 2013.

[5] S. 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, Apr. 2009.

[6] C. Chen, J. Kwon, J. Rice, A. Skabardonis, and P. Varaiya, "Detecting errors and imputing missing data for single loop surveillance systems," *Transp. Res. Rec.*, vol. 1855, pp. 160–167, 2003.

[7] C. Chui and G. Chen, *Kalman Filtering With Real-Time Applications,* 4th ed. Berlin, Germany: Springer-Verlag, 2009.

[8] S. Clark, M. Dougherty, and H. Kirby, "The use of neural networks and time series models for short-term traffic forecasting: A comparative study," in *Proc. PTRC 21st Annu. Meet.—Transportation Planning Methods*, Manchester, U.K., 1993, pp. 151–162.

[9] H. Jula, M. Dessouky, and P. Ioannou, "Real-time estimation of travel times along the arcs and arrival times at the nodes of dynamic stochastic networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 97–110, Mar. 2008.

[10] M. Karlaftis and E. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transp. Res. Part C*, vol. 19, no. 3, pp. 387–399, Jun. 2011.

[11] J. Kwon, P. Varaiya, and A. Skabardonis, "Estimation of truck traffic volume from single loop detectors with lane-to-lane speed correlation," *Transp. Res. Rec.*, no. 1856, pp. 106–117, 2003.

[12] J. Long, Z. Gao, P. Orenstein, and H. Ren, "Control strategies for dispersing incident-based traffic jams in two-way grid networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 469–481, Jun. 2012.

[13] J. McCrea and S. Moutari, "A hybrid macroscopic-based model for traffic flow in road networks," *Eur. J. Oper. Res.*, vol. 207, no. 2, pp. 676–684, Dec. 2010.

[14] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez, "A piecewise-linearized cell transmission model and parameter calibration methodology," *Transp. Res. Rec.*, no. 1965, pp. 183–191, 2006.

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

[16] T. Pan, "The stochastic dynamic journey time reliability analysis by considering the spatial and temporal correlations," M.Phil. dissertation, Dept. Civil Environmental Eng. The Hong Kong Polytechnic Univ., Hong Kong, 2012.

[17] S. Peeta and A. Ziliaskopoulos, "Foundations of dynamic traffic assignment: The past, the present, and the future," *Netw. Spatial Econom.*, vol. 1, no. 3/4, pp. 233–265, Sep. 2001.

[18] M. Schönhof and D. Helbing, "Empirical features of congested traffic states and their implications for traffic modeling," *Transp. Sci.*, vol. 41, no. 2, pp. 135–166, May 2007.

[19] J. Sheu, L. Lan, and Y. Huang, "Short-term prediction of traffic dynamics with real-time recurrent learning algorithms," *Transportmetrica*, vol. 5, no. 1, pp. 59–83, Jan. 2009.

[20] A. Sumalee, R. Zhong, T. Pan, and W. Szeto, "Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment," *Transp. Res. Part B*, vol. 45, no. 3, pp. 507–533, Mar. 2011.

[21] A. Sumalee, T. Pan, R. Zhong, N. Uno, and N. Indra-Payoong, "Dynamic stochastic journey time estimation and reliability analysis using stochastic cell transmission model: Algorithm and case studies," *Transp. Res. Part C.* [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X12001362, to be published.

[22] W. Y. Szeto, B. Ghosh, B. Basu, and M. Mahony, "Multivariate traffic forecasting technique using cell transmission model and SARIMA model," *J. Transp. Eng.*, vol. 135, no. 9, pp. 658–667, Sep. 2009.

[23] M. Tam and W. H. K. Lam, "Using automatic vehicle identification data for travel time estimation in Hong Kong," *Transportmetrica*, vol. 4, no. 3, pp. 179–194, Sep. 2008.

[24] 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.

[25] L. Tsirigotis, E. Vlahogianni, and M. Karlaftis, "Does information on weather affect the performance of short-term traffic forecasting models?" *Int. J. Intell. Transp. Syst. Res.*, vol. 10, no. 1, pp. 1–10, Jan. 2012.

[26] Y. Tong, *The Multivariate Normal Distribution*. New York, NY, USA: Spring-Verlag, 1990.

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

[28] E. Vlahogianni and M. Karlaftis, "Comparing traffic flow time-series under fine and adverse weather conditions using recurrence-based complexity measures," *Nonlinear Dyn.*, vol. 69, no. 4, pp. 1949–1963, Sep. 2012.

[29] Y. Wang and M. Papageorgiou, "Real-time freeway traffic state estimation based on extended Kalman filter: A general approach," *Transp. Res. Part B*, vol. 39, no. 2, pp. 141–167, Feb. 2005.

[30] Y. Wang, P. Coppola, A. Messmer, A. Tzimitsi, M. Papageorgiou, and A. Nuzzolo, "Real-time freeway network traffic surveillance: Large-scale field testing in Southern Italy," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 548–562, Jun. 2011.

[31] J. Yuan and K. Mills, "A cross-correlation-based method for spatial–temporal traffic analysis," *Perform. Eval.*, vol. 61, no. 2/3, pp. 163–180, Jul. 2005.

[32] R. Zhong, A. Sumalee, T. Pan, and W. H. K. Lam, "Stochastic cell transmission model for traffic network with demand and supply uncertainties," *Transportmetrica*, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X12001362, to be published.
[33] R. Zhong, T. Pan, A. Sumalee, and W. H. K. Lam, Traffic state estimation under abnormal traffic conditions using sum-of-norms regularization, Working paper.
[34] R. Zhong, A. Sumalee, T. Pan, and W. H. K. Lam, "A cell transmission model with lane changing and incorporation of stochastic demand and supply uncertainties for freeway traffic state estimation," in *Proc. 4th Int. Symp. Dyn. Traffic Assign.*, Marthas Vineyard, MA, USA, 2012.
[35] R. Zhong, A. Sumalee, T. Pan, and W. H. K. Lam, "Optimal and robust strategies for freeway traffic management under demand and supply uncertainties: An overview and general theory," *Transportmetrica*, to be published.

T. L. Pan received the B.Eng. degree in electronic engineering from Sun Yat-sen University, Guangzhou, China, in 2007 and the M.Phil. degree in transportation engineering from The Hong Kong Polytechnic University, where she is currently working towards the Ph.D. degree in transportation engineering.

Her research interests include dynamic traffic surveillance and management for intelligent transportation system applications, network modeling and optimization, network reliability, and performance assessment.

A. Sumalee received the B.Eng. degree in civil engineering from KMITL-Ladkrabang in 1999 and the M.Sc.(Eng.) and Ph.D. degrees in transportation planning and engineering from Leeds University, Leeds, U.K., in 2000 and 2004, respectively.

He is an Associate Professor with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong. His research areas include transit planning, intelligent transport systems, network modeling and optimization, and transport economics.

Dr. Sumalee is currently the Editor-in-Chief of *Transportmetica B: Transport Dynamics*, the Editor of the *Journal of Transport Policy*, an Associate Editor of *Networks and Spatial Economics*, and an Editorial Board Member of several prestigious journals.

R. X. Zhong received the B.Eng. degree in electronic engineering from Sun Yat-Sen University, Guangzhou, China, in 2005, the M.Phil. degree in automatic control from The Chinese University of Hong Kong in 2007, and the Ph.D. degree in transportation engineering from The Hong Kong Polytechnic University in 2011. He is an Associate Professor with the Research Center of Intelligent Transportation Systems, School of Engineering, Sun Yat-sen University. His main research interests include dynamic traffic surveillance, assignment, and management for intelligent transportation system applications; incident detection and management; optimization; and optimal and nonlinear control.

Dr. Zhong received the Hong Kong Society for Transportation Studies (HKSTS) Outstanding Dissertation Paper Award at the 17th HKSTS International Conference.

N. Indra-payoong received the Ph.D. degree in computing from the University of Leeds, Leeds, U.K.

He is currently an Associate Professor of logistics and supply chain management with the Faculty of Logistics, Burapha University, Bangkok, Thailand. His research interests include large-scale network optimization and intelligent transport systems.