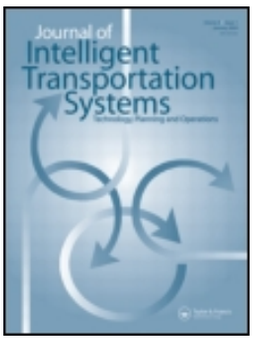

Taylor & Francis
Taylor & Francis Group

Journal of Intelligent Transportation Systems
Technology, Planning, and Operations

ISSN: 1547-2450 (Print) 1547-2442 (Online) Journal homepage: http://www.tandfonline.com/loi/gits20

Urban Traffic Flow Prediction Using a Spatio-Temporal Random Effects Model

Yao-Jan Wu, Feng Chen, Chang-Tien Lu & Shu Yang

To cite this article: Yao-Jan Wu, Feng Chen, Chang-Tien Lu & Shu Yang (2016) Urban Traffic Flow Prediction Using a Spatio-Temporal Random Effects Model, Journal of Intelligent Transportation Systems, 20:3, 282-293, DOI: 10.1080/15472450.2015.1072050

To link to this article: https://doi.org/10.1080/15472450.2015.1072050

Accepted author version posted online: 25 Jul 2015.
Published online: 11 Sep 2015.

Journal of Intelligent Transportation Systems, 20(3):282–293, 2016

ISSN: 1547-2450 print / 1547-2442 online
DOI: 10.1080/15472450.2015.1072050

Taylor & Francis
Taylor & Francis Group

Urban Traffic Flow Prediction Using a Spatio-Temporal Random Effects Model

YAO-JAN WU,[1] FENG CHEN,[2] CHANG-TIEN LU,[3] and SHU YANG[1]

[1]Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, Arizona, USA
[2]Department of Computer Science, University at Albany–SUNY, Albany, New York, USA
[3]Department of Computer Science, Virginia Tech, Falls Church, Virginia, USA

Traffic prediction is critical for the success of intelligent transportation systems (ITS). However, most spatio-temporal models suffer from high mathematical complexity and low tune-up flexibility. This article presents a novel spatio-temporal random effects (STRE) model that has a reduced computational complexity due to mathematical dimension reduction, with additional tune-up flexibility provided by a basis function capable of taking traffic patterns into account. Bellevue, WA, was selected as the model test site due to its widespread deployment of loop detectors. Data collected during the 2 weeks of July 2007 from 105 detectors in the downtown area were used in the modeling process and traffic volumes predicted for 14 detectors for the entire month of July 2008. The results show that the STRE model not only effectively predicts traffic volume but also outperforms three well-established volume prediction models, the enhanced versions of autoregressive moving average (ARMA) and spatiotemporal ARMA, and artificial neural network. Even without further model tuning, all the experimental links produced mean absolute percentage errors between 8% and 16% except for three atypical locations. Based on lessons learned, recommendations are provided for future applications and tune-up of the proposed STRE model.

Keywords Kalman Filter; Prediction Methods; Traffic Information; Traffic Operations; Traffic Prediction; Uncertainty

INTRODUCTION

Reliable, accurate, and consistent real-time traffic information is the key to success in the development and implementation of intelligent transportation systems (ITS). Subsystems of ITS, such as the Advance Traveler Information System (ATIS) and the Advance Traffic Management System (ATMS), rely heavily on high-quality real-time traffic data to provide road users with up-to-date guidance and implement traffic control schemes. In the past, the collection of real-time data was the foremost goal, but recently many agencies have begun to consider taking advantage of the vast archived data sets for "real-time forward-looking analysis" (Min & Wynter, 2011). By utilizing predicted data, proactive transportation management becomes a feasible option. For example, adaptive traffic signal control would be more effective if it were based on predicted traffic volume.

In today's ITS environment, time- and location-specific data are collected in huge volumes in real time (Cheng, Haworth, & Wang, 2011), and more and more agencies are capable of archiving these traffic data. Processing real-time and historical traffic data simultaneously could provide useful results if used in conjunction with reliable short-term traffic prediction algorithms, providing significant benefits to traffic management without the need for investment in new facilities. Unfortunately, a consistent data feed to a city's traffic management center (TMC) is not always feasible due to communication errors and malfunctioning detectors, and inconsistent data connections represent one of the key problems for arterial ATIS (Wu, An, Ma, & Wang, 2011). Maintaining consistent, high-quality traffic data flow is a challenging task for researchers and practitioners, so robust short-term traffic prediction is vital for successful ITS applications.

Travel demand forecasting also relies on accurate short-term traffic flow prediction (Smith, Williams, & Keith Oswald, 2002).

Address correspondence to Yao-Jan Wu, Department of Civil Engineering and Engineering Mechanics, University of Arizona, 1209 E. 2nd St., Tucson, AZ 85721, USA. E-mail: yaojan@email.arizona.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/gits.

Over the past three decades, research has generally focused on freeway traffic status (volume, speed, or occupancy) prediction (Smith & Demetsky, 1997; Smith, Williams, & Keith Oswald, 2002; Williams & Hoel, 2003 and references therein). Urban networks are usually more complicated than freeways, with a correspondingly greater likelihood of communication disruption, and traffic control strategies are thus frequently less responsive because of the lag between traffic data detection and implementation. Due to the complex infrastructure of urban environments such as cities, a more responsive volume prediction scheme is required, but this presents additional challenges. In terms of volume prediction method development, there are two major differences between freeways and arterials. First, the detectors are usually more closely spaced in arterial roads, and new traffic prediction methods can therefore take advantage of the geospatial relationships between detectors to provide better prediction accuracy. Second, urban traffic suffers from delays caused by signalized intersections; traffic status introduces more irregularity and uncertainty to traffic prediction because of traffic characteristics such as frequent queues and lane-changing behaviors, leading to low prediction precision on arterial networks. A new traffic prediction method consequently needs to be more responsive in order to react to rapid changes in urban traffic status. Therefore, the objective of this research is to develop an innovative traffic prediction model that can effectively and efficiently predict traffic flow for a dense urban street network with volatile traffic flows.

LITERATURE REVIEW

Despite the fact that the spatial relationships are strong and noticeable on urban networks, most research has focused on "one point" (or single detector) short-term traffic prediction that considered only the temporal domain and did not take into account the dependencies between detectors (spatial domain). These so-called "univariate" methods are similar to those used for freeway cases and are generally based on the use of time-series-based methods. The autoregressive integrated moving average (ARIMA)-based method is commonly used (see, e.g., Hamed & Al-Masaeid, 1995; Williams & Hoel, 2003). Many of the univariate models have been directly compared by researchers. Smith and Demetsky (1997) compared historical averaging, time series (ARIMA), back-propagation neural networks, and nonparametric regressions, and Smith et al. (Smith, Williams, & Keith Oswald, 2002) went on to show that the results generated by seasonal ARIMA models are statistically superior to those produced by nonparametric regression. Recent research showed that the newly developed models, such as KNN-T and the self-organizing map model, outperformed the conventional ARIMA-based model (Chiou, Lan, & Tseng, 2014; Qiao, Haghani, & Hamedi, 2013). In general, the ARIMA-based models yield satisfactory performances.

Modified univariate time-series-based approaches are still being used to predict traffic. For example, Tan et al. (Tan, Wong, Xu, & Guan, 2009) developed the data aggregation (DA) strategy to integrate moving average (MA), exponential smoothing (ES) and ARIMA models using a neural network (NN). Their proposed method shows the DA approach outperforms the naive ARIMA, nonparametric regression, and NN models. Thomas et al. (Thomas, Weijermars, & van Berkum, 2010) developed a heuristic approach to predict short- and long-term traffic. The novelty of their approach relies on their use of a mixed method that combines the concept of time series (temporal correlation) with the application of a Kalman filter to reduce the noise.

Despite these successes in single-point prediction, more and more researchers are beginning to utilize the "geographic advantage" of urban network analysis to provide better prediction results. Since arterial detectors are geographically closer to each other than freeway detectors, urban traffic prediction can rely not only on historical data, but also on real-time data from neighboring detectors (links). Therefore, in addition to its utility for short-term traffic prediction, a spatio-temporal (ST)-based predictor has a major advantage over univariate detectors in that an ST-based detector can potentially predict or estimate traffic volume simply based on data from neighboring detectors.

Recent research has therefore begun to take into account spatial information in order to improve prediction accuracy. Among all the various methods, multivariate time series such as the spatio-temporal (ST) ARIMA (Kamarianakis, Kanas, & Prastacos, 2005; Kamarianakis & Prastacos, 2003; Stathopoulos & Karlaftis, 2003), multivariate structural time series (Chandra & Al-Deek, 2009; Ghosh, Basu, & O'Mahony, 2009), dynamic STARIMA (Min, Hu, Chen, Zhang, & Zhang, 2009), and generalized STARIMA (Min, Hu, & Zhang, 2010) have been the most popular. However, time-series models involve the calibration of many parameters. Smith et al. (Smith, Williams, & Keith Oswald, 2002) compared several parametric and nonparametric traffic prediction models and found the ARIMA model to be relatively time-consuming. Due to the nature of multivariate time series, adding even one more dimension (spatial) greatly increases the computational complexity and estimation of a large number of parameters (Ghosh, Basu, & O'Mahony, 2009).

Spatio-temporal correlations have begun to attract more attentions and have recently been used to forecast traffic flow. Vlahogianni et al. (Vlahogianni, Karlaftis, & Golias, 2007) developed a traffic volume predictor that uses "temporal structures of feed-forward multilayer perceptrons (MLP)." Vlahogianni (Vlahogianni, 2009) then further enhanced this pattern-based neural network prediction scheme by considering a range of traffic flow regimes. Zou et al. (Zou, Yue, Li, & Shi, 2010) used a spatial autocorrelation method to estimate the patterns of traffic states among urban streets based on historical travel time data, although traffic flow prediction was not the primary focus of this research. Cheng et al. (Cheng, Haworth, & Wang, 2011) investigated the autocorrelation of space-time observations of traffic to determine "likely requirements for building a suitable space–time forecasting model." Most recently, a multivariate spatio-temporal autoregressive (MSTAR) model (Min & Wynter, 2011) has been designed that minimizes the number of

parameters, thus reducing the computational costs sufficiently to allow the model to be applied to large metropolitan areas. The model was tested on a large urban network.

Based on this literature review, the common challenge facing all spatio-temporal model-related research is the dimension of the network. Once the network grows beyond a certain size, most spatio-temporal models are no longer capable of handling the data generated in a timely manner. Huge data sets collected from a large network are becoming more and more common with the rapid development and deployment of ITS sensors, and this large number of spatial detectors inevitably results in a high-dimensional statistical model. To deal with this issue, a spatio-temporal random effects (STRE) model is proposed in this study.

METHODOLOGY

Inheriting the filtering capability of the Kalman filter (KF), the spatio-temporal Kalman filter (STKF) expands KF to the spatio-temporal domain. However, the traditional STKF suffers from a low performance in modeling high-dimension data (Cressie, Shi, & Kang, 2010). The STRE model, a special type of STKF proposed by Cressie et al., has proven its mathematical effectiveness in dimension reduction and parameter estimation (Cressie, Shi, & Kang, 2010). To explain the innovative STRE model used in this study, the spatial random effects (SRE) model is first introduced. After adding a temporal component, the SRE model becomes a spatio-temporal random effect (STRE) model. The details of the STRE model, in particular its parameter estimation and prediction process, will be also presented.

Spatial Random Effects (SRE) Model

Let $\left(\mathrm{Y}\left(\mathrm{s}\right):\mathrm{s}\in\mathrm{D}\in\mathbb{R}^2\right)$ be a real-valued spatial process. The spatial random effects (SRE) model first decomposes the spatial process into two additive components:

$$\mathrm{Z}\left(\mathbf{s}\right)=\mathbf{Y}\left(\mathbf{s}\right)+\in\left(\mathbf{s}\right),\mathbf{s}\in\mathrm{D} \tag{1}$$

where $\varepsilon\left(\mathrm{s}\right)$ is a spatial white process with mean zero, $\mathrm{var}\left(\varepsilon_{\mathrm{t}}\left(\mathrm{s}\right)\right)=\sigma^2_{\varepsilon,\mathrm{t}}\mathrm{v}\left(\mathrm{s}\right)>0$, $\sigma^2_{\varepsilon,\mathrm{t}}$ is a parameter to be estimated, and $\mathrm{v}\left(\mathrm{s}\right)$ is known. The white noise assumption implies that $\mathrm{cov}\left(\varepsilon\left(\mathrm{s}\right),\varepsilon\left(\mathrm{r}\right)\right)=0$, unless $\mathrm{s}=\mathrm{r}$.

The hidden process $\mathrm{Y}\left(\mathrm{s}\right)$ is assumed to have the linear mean structure:

$$\mathrm{Y}\left(\mathbf{s}\right)=\mathbf{x}\left(\mathbf{s}\right)^{\mathrm{T}}\boldsymbol{\beta}+\upsilon\left(\mathbf{s}\right),\mathbf{s}\in\mathrm{D} \tag{2}$$

where $\mathrm{x}\left(\mathrm{s}\right)$ is a vector of known covariates, the coefficients β are unknown, and the process $\upsilon\left(\mathrm{s}\right)$ is a spatial process with zero mean and a general nonstationary spatial covariance function that is captured by a set of basis functions $\{\mathrm{b}_1\left(\mathrm{s}\right),\ldots,\mathrm{b}_{\mathrm{r}}\left(\mathrm{s}\right)\}$ as,

$$\upsilon\left(\mathbf{s}\right)=\mathbf{b}\left(\mathbf{s}\right)^{\mathrm{T}}\boldsymbol{\eta}+\xi\left(\mathbf{s}\right) \tag{3}$$

where $\mathrm{b}\left(\mathrm{s}\right)=\left[\mathrm{b}_1\left(\mathrm{s}\right),\ldots,\mathrm{b}_{\mathrm{p}}\left(\mathrm{s}\right)\right]^{\mathrm{T}}$, η is a vector of r-dimensional Gaussion process with mean zero and covariances K: $\eta\sim\mathcal{N}_{\mathrm{r}}\left(0,\mathrm{K}\right)$, and $\xi\left(\mathrm{s}\right)$ is independent Gaussian white noise with zero mean and variance σ^2_{ξ}. Then, by combining Eqs. 1, 2, and 3, we have the SRE model as:

$$\mathrm{Z}\left(\mathbf{s}\right)=\mathbf{x}\left(\mathbf{s}\right)^{\mathrm{T}}\boldsymbol{\beta}+\mathbf{b}\left(\mathbf{s}\right)^{\mathrm{T}}\boldsymbol{\eta}+\varepsilon\left(\mathbf{s}\right)+\xi\left(\mathbf{s}\right) \tag{4}$$

The unknown parameters are $\left\{\beta,\sigma^2_{\varepsilon,\mathrm{t}},\sigma^2_{\xi}\right\}$. By employing this form, it can be shown that the resulting best linear unbiased predictor (BLUP) achieves significant computational savings compared with a traditional Kriging model. The BLUP estimator based on the SRE model is equivalent to a fixed-rank Kriging.

Spatio-Temporal Random Effects Model (STRE)

The spatio-temporal random effects (STRE) model is an extension of the SRE model that incorporates a consideration of temporal effects. The STRE model can perform the following tasks: dimension reduction (spatial) and rapid smoothing, filtering, or prediction (temporal) (Cressie, Shi, & Kang, 2010). Filtering, smoothing, and prediction based on STRE are also known as fixed rank filtering (FRF), fixed rank smoothing, and fixed rank prediction (Cressie, Shi, & Kang, 2010).

The STRE model can be used to model a spatial random process that evolves over time, $\left\{\mathrm{Y}_t\left(\mathbf{s}\right)\in\mathbb{R}:\mathbf{s}\in\mathrm{D}\in\mathbb{R}^2,t=1,2,\ldots\right\}$, where D is the spatial domain under study, and $\mathrm{Y}_{\mathrm{t}}\left(\mathbf{s}\right)$ represents the measurements at location $\mathbf{s}$ and time t. A discretized version of the process can be represented as:

$$\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_t,\mathbf{Y}_{t+1},\ldots \tag{5}$$

where $\mathbf{Y}_t=\left[\mathrm{Y}_t\left(\mathbf{s}_{1,t}\right),\mathrm{Y}_t\left(\mathbf{s}_{2,t}\right),\ldots,\mathrm{Y}_t\left(\mathbf{s}_{m_t,t}\right)\right]^{\mathrm{T}}$. The sample locations $\left\{\mathbf{s}_{1,t},\mathbf{s}_{2,t},\ldots,\mathbf{s}_{m_t,t}\right\}$ can be different spatial locations at different times t.

Two major uncertainties, namely, missing data and noise (measurement error), can be handled in this model. Suppose we have the measurements $\{\mathbf{Z}_1,\mathbf{Z}_2,\ldots,\}$, with:

$$\mathbf{Z}_t=\mathbf{O}_t\mathbf{Y}_t+\varepsilon_t,t=1,2,\ldots, \tag{6}$$

where $\mathbf{Z}_{\mathrm{t}}$ is an n_t-dimensional vector $(n_t\leq m_t)$, $\mathbf{O}_{\mathrm{t}}$ is an $n_t\times m_t$ incidence matrix, and $\varepsilon_t=\left[\varepsilon_t\left(\mathbf{s}_{1,t}\right),\varepsilon_t\left(\mathbf{s}_{2,t}\right),\ldots,\varepsilon_t\left(\mathbf{s}_{m_t,t}\right)\right]^{\mathrm{T}}\sim\mathcal{N}_{\mathrm{m_t}}\left(0,\sigma^2_{\varepsilon,t}\mathbf{V}_{\varepsilon,t}\right)$ is a vector of white noise Gaussian processes, with $\mathbf{V}_{\varepsilon,t}=\mathrm{diag}\left(\mathrm{v}_{\varepsilon,t}\left(\mathbf{s}_{1,t}\right),\ldots,\mathrm{v}_{\varepsilon,\mathrm{t}}\left(\mathbf{s}_{\mathrm{n_t},t}\right)\right)$. Particularly, $\mathrm{var}\left(\varepsilon_t\left(\mathbf{s}\right)\right)=\sigma^2_{\varepsilon,t}\mathrm{v}\left(\mathbf{s}\right)>0$, $\sigma^2_{\varepsilon,t}$ is a parameter to be estimated, and $\mathrm{v}\left(\mathbf{s}\right)$ is known. The white noise assumption implies that $\mathrm{cov}\left(\varepsilon_t\left(\mathbf{s}\right),\varepsilon_u\left(\mathbf{r}\right)\right)=0$, for $\mathrm{t}\neq\mathrm{u}$ and $\mathbf{s}\neq\mathbf{r}$.

Assume that $\mathbf{Y}_{\mathrm{t}}$ has the following structure:

$$\mathbf{Y}_t=\boldsymbol{\mu}_t+\boldsymbol{\upsilon}_t,t=1,2,\ldots, \tag{7}$$

where $\boldsymbol{\mu}_t$ is a vector of deterministic (spatio-temporal) mean or trend functions modeling large-scale variations, and the random process $\boldsymbol{\upsilon}_t$ captures the small-scale variations. A

common strategy is to define $\boldsymbol{\mu}_t = \mathbf{X}_t\boldsymbol{\beta}_t$, where $\mathbf{X}_t = \left[\mathbf{x}_t\left(\mathbf{s}_{1,t}\right), \ldots, \mathbf{x}_t\left(\mathbf{s}_{n_t,t}\right)\right]^{\mathrm{T}}$ and $\mathbf{x}_t\left(\mathbf{s}_{1,t}\right) \in^{\mathrm{p}}$ represents a vector of covariates. The coefficients $\boldsymbol{\beta}_t$ are in general unknown and need to be estimated or predefined.

In challenging applications such as astronomy, the values n_t and m_t can be large-scale in nature. For traditional spatio-temporal Kalman filtering models, a large number of parameters need to be estimated, with the accompanying high computational costs due to the high data dimensionality during the filtering, smoothing, and prediction processes. A key advantage of the STRE model is that it is capable of modeling the small-scale variation $\boldsymbol{\upsilon}_t$ as a vector of spatial random effects (SRE) processes:

$$\boldsymbol{\upsilon}_t = \mathbf{B}_t\boldsymbol{\eta}_t + \xi_t, t = 1, 2, \ldots, \tag{8}$$

where $\mathbf{B}_t = \left[\mathbf{b}_t\left(\mathbf{s}_{1,t}\right), \ldots, \mathbf{b}_t\left(\mathbf{s}_{n_t,t}\right)\right]^{\mathrm{T}}$, $\mathbf{b}_t\left(\mathbf{s}_{1,t}\right) = \left[\mathrm{b}_{1,t}\left(\mathbf{s}_{1,t}\right), \ldots, \mathrm{b}_{r,t}\left(\mathbf{s}_{1,t}\right)\right]^T$ is a vector of r predefined spatial basis functions such as wavelet and bisquare basis functions, and $\boldsymbol{\eta}_t$ is a zero-mean Gaussian random vector with an r × r covaraince matrix given by $\mathbf{K}_t$. The first component in Eq. 8 denotes a smoothed small-scale variation at time t, captured by the set of basis functions $\left\{\mathbf{b}_t\left(\mathbf{s}_{1,t}\right), \ldots, \mathbf{b}_t\left(\mathbf{s}_{n_t,t}\right)\right\}$.

The second component in Eq. 8 captures the fine-scale variability similar to the nugget effect as defined in geostatistics (Cressie, Shi, & Kang, 2010). It is assumed that $\boldsymbol{\xi}_{\mathrm{t}} \sim \mathcal{N}_{\mathrm{m_t}}\left(0, \sigma^2_{\xi,\mathrm{t}}\mathbf{V}_{\xi,\mathrm{t}}\right)$, $\mathbf{V}_{\xi,t} = \mathrm{diag}\left(\mathrm{v}_{\xi,t}\left(\mathbf{s}_{1,t}\right), \ldots, \mathrm{v}_{\xi,t}\left(\mathbf{s}_{n_t,t}\right)\right)$, and $\mathrm{v}_{\xi,t}(\cdot)$ describes the variance of the fine-scale variation and is typically considered known. If no expert knowledge is available about the variance form, it could be modeled as $\mathrm{v}_{\xi,t}(\cdot) = \exp\left\{\mathbf{b}_\xi^T\boldsymbol{\eta}_\xi\right\}$, where $\mathbf{b}_\xi$ is a vector of r_ξ basis functions, with $\mathrm{r}_\xi < r$ (Katzfuss & Cressie, 2011). Note that the component $\boldsymbol{\xi}_t$ is important, since it can be used to capture the extra uncertainty due to the dimension reduction in replacing $\boldsymbol{\upsilon}_t$ by $\mathbf{B}_t\boldsymbol{\eta}_t$. The coefficient vector $\boldsymbol{\eta}_t$ is assumed to follow a vector-autoregressive process of order 1,

$$\boldsymbol{\eta}_t = \mathbf{H}_t\boldsymbol{\eta}_{t-1} + \boldsymbol{\zeta}_t, t = 1, 2, \ldots, \tag{9}$$

where $\mathbf{H}_t$ refers to the so-called propagator matrix, $\boldsymbol{\zeta}_t \sim \mathcal{N}(0, \mathbf{U}_t)$ is an r-dimensional innovation vector, and $\mathbf{U}_t$ is referred to as the innovation matrix. The initial state $\boldsymbol{\eta}_0 \sim \mathcal{N}_{\mathrm{r}}(0, \mathbf{K}_0)$ and K_0 are in general unknown.

Combining Eqs. 6, 7, and 8, the (discretized) data process can be represented as:

$$\mathbf{Z}_t = \mathbf{O}_t\boldsymbol{\mu}_t + \mathbf{O}_t\mathbf{B}_t\boldsymbol{\eta}_t + \mathbf{O}_t\boldsymbol{\xi}_t + \boldsymbol{\varepsilon}_{\mathrm{t}}, t = 1, \ldots, \tag{10}$$

where $\boldsymbol{\mu}_{\mathrm{t}}$ is deterministic and the other components are stochastic (Cressie, Shi, & Kang, 2010).

Filtering, Smoothing, and Prediction

The STRE model can perform filtering, smoothing, and prediction, with the relevant mathematical operations defined as follows: Let $\boldsymbol{\eta}_{t|\hat{t}} = \mathrm{E}\left(\boldsymbol{\eta}_t|\mathbf{z}_{1:\hat{t}}\right)$, $\boldsymbol{\xi}_{t|\hat{t}} = \mathrm{E}\left(\boldsymbol{\xi}_t|\mathbf{z}_{1:\hat{t}}\right)$. Denote $\mathbf{P}_{t|\hat{t}} = \mathrm{var}\left(\boldsymbol{\eta}_t|\mathbf{z}_{1:\hat{t}}\right)$ as the conditional covariance matrix of $\boldsymbol{\eta}_t$, and $\mathbf{R}_{t|\hat{t}} = \mathrm{var}\left(\boldsymbol{\xi}_t|\mathbf{z}_{1:\hat{t}}\right)$ as the conditional covariance matrix $\boldsymbol{\xi}_t$. For the initial state, we set $\boldsymbol{\eta}_{0|0} = 0$ and $\mathbf{P}_{0|0} = \mathbf{K}_0$.

The fixed rank filtering estimator of $\mathbf{Y}_t$ is:

$$\begin{aligned}
\mathbf{Y}_{t|t} &= \boldsymbol{\mu}_t + \mathbf{B}_t\boldsymbol{\eta}_{t|t} + \boldsymbol{\xi}_{t|t} \\
\boldsymbol{\eta}_{t|t} &= \boldsymbol{\eta}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{B}_t^{\mathrm{T}}\mathbf{O}_t^{\mathrm{T}}\left[\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1}\mathbf{B}_t^{\mathrm{T}}\mathbf{O}_t^{\mathrm{T}} + \mathbf{D}_t\right]^{-1} \\
&\quad \times \left(\mathbf{z}_t - \mathbf{O}_t\mathbf{X}_t\boldsymbol{\beta}_t - \mathbf{O}_t\mathbf{B}_t\boldsymbol{\eta}_{t|t-1}\right), \\
\boldsymbol{\xi}_{t|t} &= \sigma^2_{\xi,t}\mathbf{V}_{\xi,t}\mathbf{O}_t^{\mathrm{T}}\left[\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1}\mathbf{B}_t^{\mathrm{T}}\mathbf{O}_t^{\mathrm{T}} + \mathbf{D}_t\right]^{-1} \\
&\quad \times \left(\mathbf{z}_t - \mathbf{O}_t\mathbf{X}_t\boldsymbol{\beta}_t - \mathbf{O}_t\mathbf{B}_t\boldsymbol{\eta}_{t|t-1}\right), \\
\mathbf{P}_{t|t} &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{B}_t^{\mathrm{T}}\mathbf{O}_t^{\mathrm{T}} \\
&\quad \times \left[\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1}\mathbf{B}_t^{\mathrm{T}}\mathbf{O}_t^{\mathrm{T}} + \mathbf{D}_t\right]^{-1}\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1}, \\
\mathbf{R}_{t|t} &= \sigma^2_{\xi,t}\mathbf{V}_{\xi,t} - \sigma^2_{\xi,t}\mathbf{V}_{\xi,t}\mathbf{O}_t^{\mathrm{T}} \\
&\quad \times \left[\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1}\mathbf{B}_t^{\mathrm{T}}\mathbf{O}_t^{\mathrm{T}} + \mathbf{D}_t\right]^{-1}\mathbf{O}_t\mathbf{V}_{\xi,t}\sigma^2_{\xi,t}
\end{aligned} \tag{11}$$

where $\mathbf{D}_t = \sigma^2_{\xi,t}\mathbf{O}_{\mathrm{t}}\mathbf{V}_{\xi,t}\mathbf{O}_t^{\mathrm{T}} + \sigma^2_{\xi,t}\mathbf{V}_{\xi,t}$

The fixed rank smoothing estimator of $\mathbf{Y}_t$, $\mathrm{t} \in \{1, 2, \ldots, T\}$, is:

$$\begin{aligned}
\mathbf{Y}_{t|T} &= \boldsymbol{\mu}_t + \mathbf{B}_t\boldsymbol{\eta}_{t|T} + \boldsymbol{\xi}_{t|T} \\
\boldsymbol{\eta}_{t|T} &= \boldsymbol{\eta}_{t|t} + \mathbf{J}_{\mathrm{t}}\left(\boldsymbol{\eta}_{t+1|T} - \boldsymbol{\eta}_{t+1|t}\right) \\
\boldsymbol{\xi}_{t|T} &= \boldsymbol{\xi}_{t|t} - \mathbf{M}_{\mathrm{t}}\left(\boldsymbol{\eta}_{t+1|T} - \boldsymbol{\eta}_{t+1|t}\right) \\
\mathbf{P}_{t|T} &= \mathbf{P}_{t|t} + \mathbf{J}_t\left(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}\right)\mathrm{J}_t^{\mathrm{T}} \\
\mathbf{R}_{t|T} &= \mathbf{R}_{t|t} + \mathbf{M}_t\left(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t}\right)\mathbf{M}_t^{\mathrm{T}}
\end{aligned} \tag{12}$$

where $\mathbf{J}_t = \mathbf{P}_{t|t}\mathbf{H}_{t+1}^{\mathrm{T}}\mathbf{P}_{t+1|t}^{-1}$, $\mathbf{M}_t = \sigma^2_{\xi,t}\mathbf{V}_{\xi,t}\mathbf{O}_t^{\mathrm{T}}\left[\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1}\mathbf{B}_t^{\mathrm{T}}\mathbf{O}_t^{\mathrm{T}} + \mathbf{D}_t\right]^{-1}\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1}\mathbf{H}_{t+1}^{\mathrm{T}}\mathbf{P}_{t+1|t}^{-1}$.

The fixed rank prediction estimator of $\mathbf{Y}_u$, $u \in \{t+1, t+2, \ldots\}$, is:

$$\begin{aligned}
\mathbf{Y}_{\mathrm{u|t}} &= \mathbf{O}_{\mathrm{u}}\boldsymbol{\mu}_{\mathrm{u}} + \mathbf{O}_{\mathrm{u}}\mathbf{B}_{\mathrm{u}}\boldsymbol{\eta}_{\mathrm{u|t}} \\
\mathbf{P}_{\mathrm{u|t}} &= \left(\prod_{\mathrm{i=t+1}}^{\mathrm{u}}\mathbf{H}_{\mathrm{i}}\right)\mathbf{P}_{\mathrm{t|t}}\left(\prod_{\mathrm{i=t+1}}^{\mathrm{u}}\mathbf{H}_{\mathrm{i}}\right)^{\mathrm{T}} + \mathbf{U}_{\mathrm{u}} \\
&\quad + \sum_{\mathrm{i=t+1}}^{\mathrm{u-1}}\left\{\left(\prod_{\mathrm{j=i+1}}^{\mathrm{u}}\mathbf{H}_{\mathrm{j}}\right)\mathbf{U}_{\mathrm{i}}\left(\prod_{\mathrm{j=i+1}}^{\mathrm{u}}\mathbf{H}_{\mathrm{j}}\right)^{\mathrm{T}}\right\}
\end{aligned} \tag{13}$$

where fixed rank filtering is aimed at estimating the traffic speeds of roadway links in historical time intervals that have no sensors, given the speed observations of other roadway links in historical time intervals and the present time intervals; fixed rank smooth-

ing is aimed at estimating the traffic speeds of roadway links that have no sensors at the present time interval, given the speed observations of other roadway links in historical time intervals and the current time interval; and fixed rank prediction is aimed at predicting the traffic speeds of roadway links in future time intervals.

Computational Complexity

The computational complexity is calculated based on the total number of observed time stamps, the total number of observed spatial locations n_t at time t, and the number of bases used in the hidden process $\{\eta_t\}$. We can compare the computational complexity between the traditional spatio-temporal Kalman filtering (STKF) model (Cressie & Wikle, 2002) and the STRE model. Given a set of observed data $\{z_1, \ldots, z_t\}$, the computational complexity of the spatio-temporal Kalman filtering is $O(\sum_t n_t^3)$, while the computational complexity of the fixed-rank filtering based on the STRE model is $O(\sum_t n_t r^3)$. In general, r is fixed with $r \ll n$. Therefore, we have the computational complexity for the STRE model as $O(\sum_t n_t)$, which is linear in order for complexity and thus indicates that the STRE model achieves significant computational savings compared with traditional STKF. For the model parameter estimation process, the expectation-maximization (EM) algorithm (Katzfuss & Cressie, 2011) is used.

DATA

The traffic volume data were collected in the city of Bellevue in Washington state, using advance loop detectors located 30.5–39.7 m (100–130 ft) upstream from the stop bar in each case. As of July 2010, the city had more than 182 signalized intersections, 165 of which are controlled by the traffic management center (TMC). Data from 706 loop detectors are sent to the TMC every minute. The data is currently managed by the Digital Roadway Interactive Visualization and Evaluation (DRIVE) Net system (Ma, Wu, & Wang, 2011; Wu, An, Ma, & Wang, 2011), at the Smart Transportation Application and Research Laboratory (STAR Lab) at the University of Washington (UW), Seattle, WA.

This study focused specifically on the downtown area because the intersections are closer to each other (around 500 feet (152.5 m) apart). The STRE model is expected to benefit from the high correlation between detectors due to this proximity of intersections. The downtown area in Figure 1 was selected as the test site, and in total 105 detectors in this area are included in the modeling process. 8th Ave was selected as the test route because it is a relatively busy street, with an annual average weekday traffic of 37,700 (vehicles/day), connecting freeway I-405 with a large shopping mall (Bellevue Square). Traffic on 8th Ave is fairly volatile and poses a challenging case study with which to examine the model's capability. Fourteen detectors, seven eastbound and seven westbound, on 8th Ave were used in the experiment. Since each link has only one detector, a link also represents a detector hereafter in this study. The links and reference points used in this study are illustrated in Figure 1; each reference point is overlapped with the intersection number, and the relevant concepts are explained in the next section.

Weekday data (Tuesday, Wednesday, and Thursday) collected from the first 2 weeks of June 2007 are used for training, and the last 2 weeks of June 2007 are used for cross-validation in this study. The verification data consist of those collected during the entire month of July 2008. All data are aggregated into 5-min intervals to reduce the effect of random noise. It should be noted that the cross-validation process is only applied to the traditional prediction methods that will be compared with the STRE model; the STRE model does not require cross-validation because the model is a variant of the first-order autoregressive model, in which the order is set to 1 as a fixed value. The other parameters of STRE are estimated based on the time-series data. Due to the large number of parameters involved, it was not considered practical to cross-validate all possible combinations of the parameters.

MODEL ADJUSTMENT

Basis Function

Before the modeling process can commence, the basis functions in Eq. 8 must be determined. For the selection of basis function, the bisquare function is used here and is defined as:

$$b_i(\mathbf{s}) = \left\{1 - \frac{\mathbf{s} - \mathbf{c}}{w}\right\}^2 I(\mathbf{s} - \mathbf{c} < w) \tag{14}$$

where c is the reference point, w is the range, and $I(\cdot)$ is an indicator function.

The range parameter determines the degree of independency between two links. The smaller the range parameter is, the more likely the two links are to be independent. Based on the experimental results, the east–west distance between downtown boundaries was selected as being most suitable for our study. It should be noted that the basis function may be dependent on time of week and time of day.

Since the basis function determines the portion of how much predicted volume each reference point should contribute depending on the correlation between the detector and each reference point, the location and number of the reference points are critical determinants of prediction accuracy. Setting the reference point as the point between two detectors and since the detectors are assumed to be in the middle of the link, each reference point is therefore located at the intersection (node) in this study. As our results show, the higher the number of reference points included in the analysis, the better the results will be, although computational efficiency will decrease. In order to increase model performance, the number of center points needs to be relatively small. Here, 11 reference points are

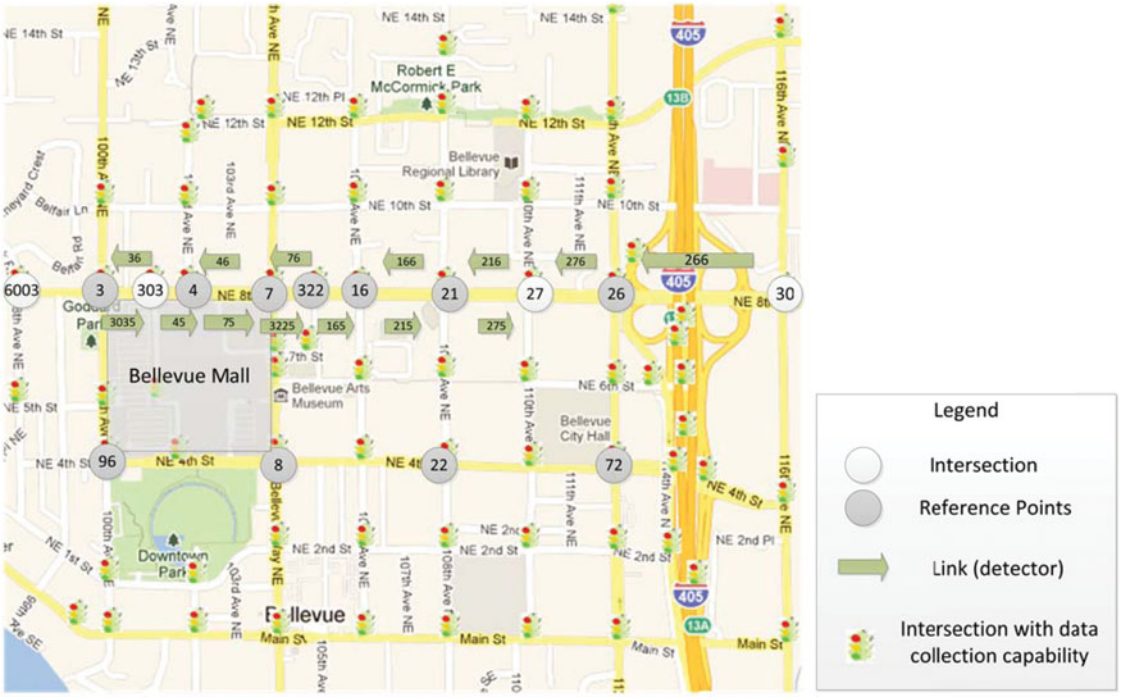

Figure 1 Downtown area in the city of Bellevue, WA (background images are from maps.google.com).

considered, illustrated in Figure 1 as dark circles. It should be noted that, unlike regular spatio-temporal data, the data collected in a transportation network must consider the direction of traffic flow. Two links with reversed directions between same pair of intersections would thus overlap with each other, and the reference points determined by these two pairs of links will also overlap (at the same intersection), but with opposite directions.

Directional Penalty

Generally, spatio-temporal models simply consider the distances between data observation points to determine correlations. In traffic applications, the L1 distance (Manhattan distance) is more reasonable than Euclidean distance and is therefore used to calculate the spatial distance between detectors. The correlation between different detectors depends not only on their spatial distance, but also, importantly, on their traffic directions and traffic turning movement counts. In order to take all these factors into account, a penalty value, p, needs to be assigned to each basis function. The revised basis function is reformulated as:

$$b_i(\mathbf{s})^* = p * b_i(\mathbf{s}) \tag{15}$$

Note that the greater the penalty value (basis function), the lower is the correlation between the reference point and the detector: The detector contributes less volume to the reference point. According to this principle, all the p values are first assigned numbers sequentially, for example, 1, 2, 3 . . ., according to the relative relationships between the detector and the reference point. Many trial-and-error operations are then conducted to determine the most suitable values, which constitutes one of the tune-up processes for the model.

Consider the intersection in Figure 2. To determine the appropriate penalty for the contribution of a detector to the reference point for traffic traveling in an eastbound direction, the rules are defined as follows:

Rule 1: If the detector is upstream of the reference point, then we use penalty $p = 1$; that is, Detector 1 contributes most of the volume to the reference point.

Rule 2: Similar to Rule 1, but the detector is downstream of the reference point. Then, penalty $p = 1.2$; that is, Detector 2 has a reduced volume contribution to the reference point.

Rule 3: If the detector direction and the reference point direction are opposite, then the penalty is set as 0, meaning their correlation is not considered; that is, Detector 3 has no contribution to the reference point because it is assumed that U-turn traffic is insignificant.

Rule 4: If the detector direction and the reference point direction are perpendicular, the penalty p is set as 7.5; that is, Detector 4 or Detector 5 makes only minor contributions to the reference point. This is because the traffic detected by Detector 1 is less likely to also be collected by Detector 4 or Detector 5, since only through-traffic detectors are used in this study.

Note that all these penalties have been determined as a result of a number of trial-and-error operations.

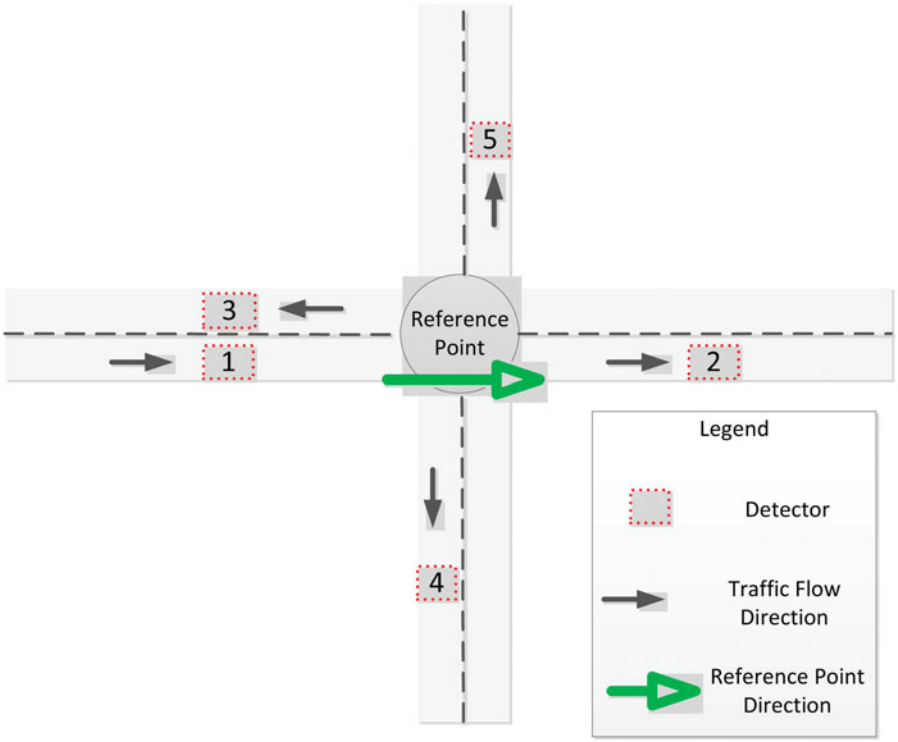

Figure 2 Basis function penalty assignment.

PREDICTION PERFORMANCE

In order to verify the STRE model performance, two measures of effectiveness are used: mean absolute percentage error (MAPE) and root mean square error (RMSE). These two measures are widely used to evaluate traffic prediction performance (Tan, Wong, Xu, Guan, & Zhang, 2009; Xie, Zhang, & Ye, 2007; Washington, Karlaftis, & Mannering, 2003) and are defined as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{G_t - Z_t}{G_t}\right| \tag{16}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(G_t - Z_t)^2}{n}} \tag{17}$$

where G_t refers to the predicted value at time t and Z_t refers to the true observation at time t.

EXPERIMENTAL RESULTS

Prediction Evaluation

To evaluate the temporal transferability of the STRE model, the model was verified with the data collected during the entire month of July 2008 (Tuesday, Wednesday, and Thursday). Only peak hours (6 a.m.–9 p.m., 180 intervals of 5 min) are considered because traffic variations are more dramatic during this period and can thus provide an effective test of the robustness of different traffic volume prediction models. In our analysis, the prediction results of all the links are separated into two groups, namely, the mall and non-mall areas, because these two areas have different traffic patterns. The reference point, 322, was designated the boundary separating the two distinct traffic patterns. The mall area (Bellevue Square) has about 180 retail stores and more than 10,000 parking spaces, attracting more than 43,000 visitors daily, so the parking lots around the mall are likely to create irregular traffic patterns that disturb the spatio-temporal prediction accuracy.

Table 1 shows the model verification results divided into two areas and using scenarios based on 1-, 5-, 15-, and 60-step prediction horizons. Here, the k-step prediction horizon means that the traffic volume predicted at $t + 60$ is 60 steps ahead of the current time t with all the traffic data missing between t and $t + k$. As expected, the prediction accuracy degrades as the prediction horizon increases, although the prediction accuracy is only slightly degraded because the proposed STRE model can take the advantage of the "adjacent data" between t and $t + k$ to estimate the unknown volume at time stamp $t \pm k$. Overall, the prediction results are satisfactory. Figure 3α and 3b show the results for Links 165 and 36, respectively, revealing two distinct patterns in the downtown area and demonstrating the challenges inherent in these datasets. This result shows that the STRE model is capable of adapting to many traffic patterns.

In terms of the prediction accuracy for different areas, the prediction MAPEs in the non-mall area are between 11.3% (1-step) and 12.5% (60-step), while the MAPEs in the mall area

Table 1 Prediction results.

Prediction horizon	Link NO	Mall area							Non-mall area						
		Eastbound				Westbound			Eastbound			Westbound			
		3035	45	75	3225	76	46	36	165	215	275	266	276	216	166
1-step	RMSE	77.178	147.619	88.318	77.548	81.072	80.817	49.361	82.851	121.965	99.742	62.498	80.160	90.661	83.526
	MAPE	0.152	0.282	0.192	0.165	0.124	0.113	0.153	0.107	0.181	0.127	0.084	0.103	0.093	0.093
	Avg. MAPE (1)			0.198			0.130			0.138				0.093	
	Avg. MAPE (2)				0.169							0.113			
5-step	RMSE	81.192	159.640	97.016	83.778	87.737	86.467	51.171	90.218	132.171	105.779	67.328	87.289	99.681	91.914
	MAPE	0.160	0.306	0.212	0.179	0.134	0.121	0.159	0.116	0.197	0.135	0.091	0.113	0.103	0.103
	Avg. MAPE (1)			0.214			0.138			0.149				0.152	
	Avg. MAPE (2)				0.181							0.152			
15-step	RMSE	81.320	161.521	98.688	84.247	88.477	86.924	51.219	91.078	135.059	107.341	68.139	89.119	101.164	93.255
	MAPE	0.160	0.310	0.215	0.180	0.135	0.121	0.159	0.117	0.202	0.137	0.093	0.116	0.104	0.104
	Avg. MAPE (1)			0.216			0.138			0.152				0.104	
	Avg. MAPE (2)				0.183							0.125			
60-step	RMSE	81.328	161.746	98.888	84.269	88.507	86.953	51.221	91.148	135.311	107.482	68.177	89.206	101.205	93.298
	MAPE	0.160	0.311	0.216	0.180	0.135	0.121	0.159	0.118	0.203	0.137	0.093	0.116	0.104	0.104
	Avg. MAPE (1)			0.217			0.138			0.152				0.104	
	Avg. MAPE (2)				0.183							0.125			

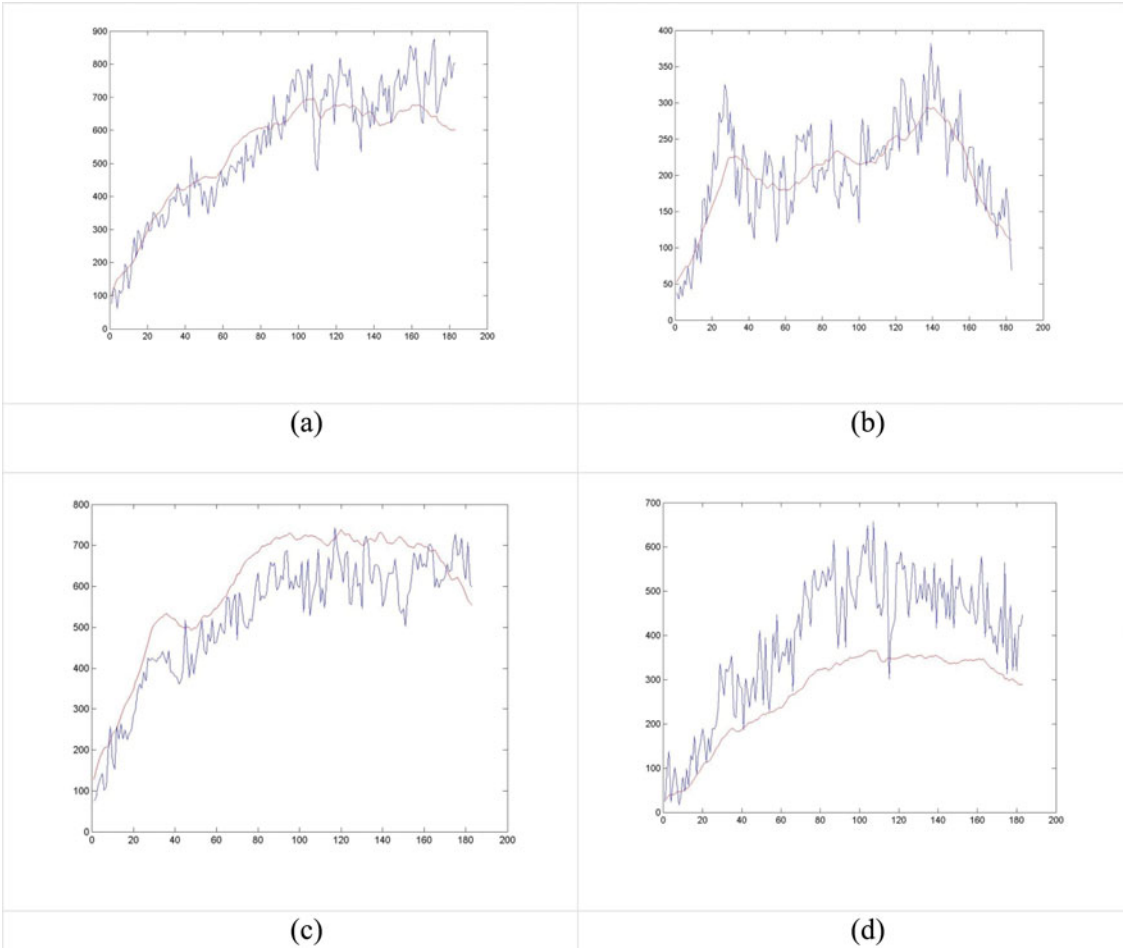

Figure 3 Examples of prediction results: (a) July 6, 2008 (Wednesday) (Link 165), and (b) July 5, 2008 (Tuesday) (Link 36). Links with low prediction accuracy: (c) July 5, 2008 (Tuesday) (Link 215), (d) July 5, 2008 (Tuesday) (Link 45) (y-axis: traffic flow rate [vehicles/hour], x-axis: time intervals).

are between 16.9% and 18.3%; the resulting RMSEs follow the same trend. It should be noted that the RMSEs diverge insignificantly, indicating that the STRE model is effective in reducing prediction randomness as the prediction horizon increases.

In the non-mall area (westbound), the overall prediction accuracy is also satisfactory (average MAPE = 11.3%). Traffic signals are coordinated for westbound traffic, and westbound traffic thus progresses more smoothly. In contrast, the STRE model tends to overestimate the volume on Link 215 (average MAPE = 19.6%), as shown in Figure 3c. This link represents a special data collection point where the city estimates the volume based on readings provided by upstream and downstream detectors, so the ground-truth data being used are actually estimated values. For the links in the mall area, the predictions for Link 45 exhibit the lowest performance (average MAPE = 30%) among all predictions. This result is not surprising because the link is located at the major entrance and exit of the parking lot, where the traffic pattern is fairly unstable. As shown in Figure 3d, the STRE model tends to underestimate the volume because the model may be unable to capture the random volume in the mall parking lot.

For both the mall and non-mall areas, westbound traffic predictions are generally better than for the eastbound predictions. This is probably because the traffic control coordination system is designed to favor westbound traffic. A semi-actuated coordinated signal control scheme is implemented on 8th Ave to release traffic from the off ramp on the freeway (I-405). This finding suggests that the implementation of traffic control schemes could be considered in future model tuning processes.

Comparisons Between Predicted and Actual Traffic Volume

To investigate the prediction performance based on different prediction horizons, a predicted vs. actual traffic volume plot can provide useful information. Figure 4 compares the actual and predicted volumes at Links 45 and 266 because these two links represent two extreme cases, namely the best and worst prediction results, as indicated in Table 1. Only two days of data are plotted to avoid cluttered data visualization. Figure 4a illustrates how the STRE model underestimates traffic volume at Link 45 due to the unpredictable traffic volume produced by a parking lot. Link 45 has the worst prediction results in our research with an average MAPE = 30%. As the volume increases and the prediction horizon extends further out, this underestimation becomes more severe. In contrast, Figure 4b shows the STRE model prediction performance at Link 266, one of the locations with the best results. This link has an average MAPE of only 9% across all prediction horizons. All the data points are close to the 45-degree line, showing the effectiveness of STRE prediction for this link regardless of prediction horizons.

Comparisons of Model Prediction Performance

To evaluate model prediction performance, the STRE model was also compared with three other models: autoregressive moving average (ARMA), spatiotemporal autoregressive moving average (STARMA) and an artificial neural network (ANN), all of which are feedforward network models (MathWorks, 2013). ARMA is a traditional prediction model that does not take into account spatiotemporal effects, whereas STARMA does consider data from adjacent traffic detectors. These well-established models have all been applied and tested in many previous volume prediction research projects (Kamarianakis & Prastacos, 2003; Min, Hu, Chen, Zhang, & Zhang, 2009; Min, Hu, & Zhang, 2010). It should be noted that our application used enhanced versions of the autoregressive moving average (ARMA) and spatiotemporal autoregressive moving average (STARMA) by Min et al. (Min, Hu, & Zhang, 2010) where both ARMA and STARMA utilize the delete-average operation rather than the difference operation to stabilize traffic flow time series, which permits the modified STARMA to process nonstationary time series. It was found that the enhanced versions of the ARMA

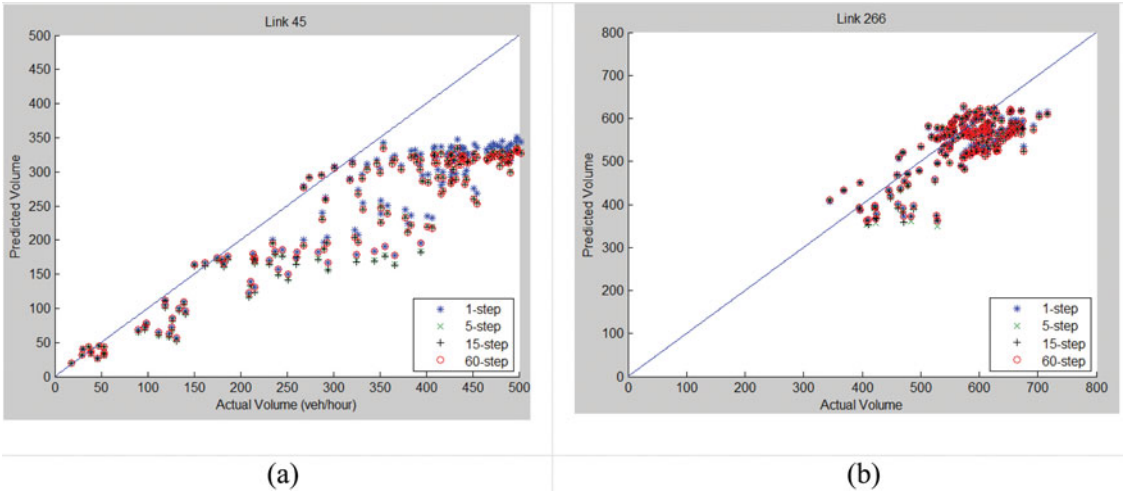

Figure 4 Comparisons between predicted and actual traffic volumes.

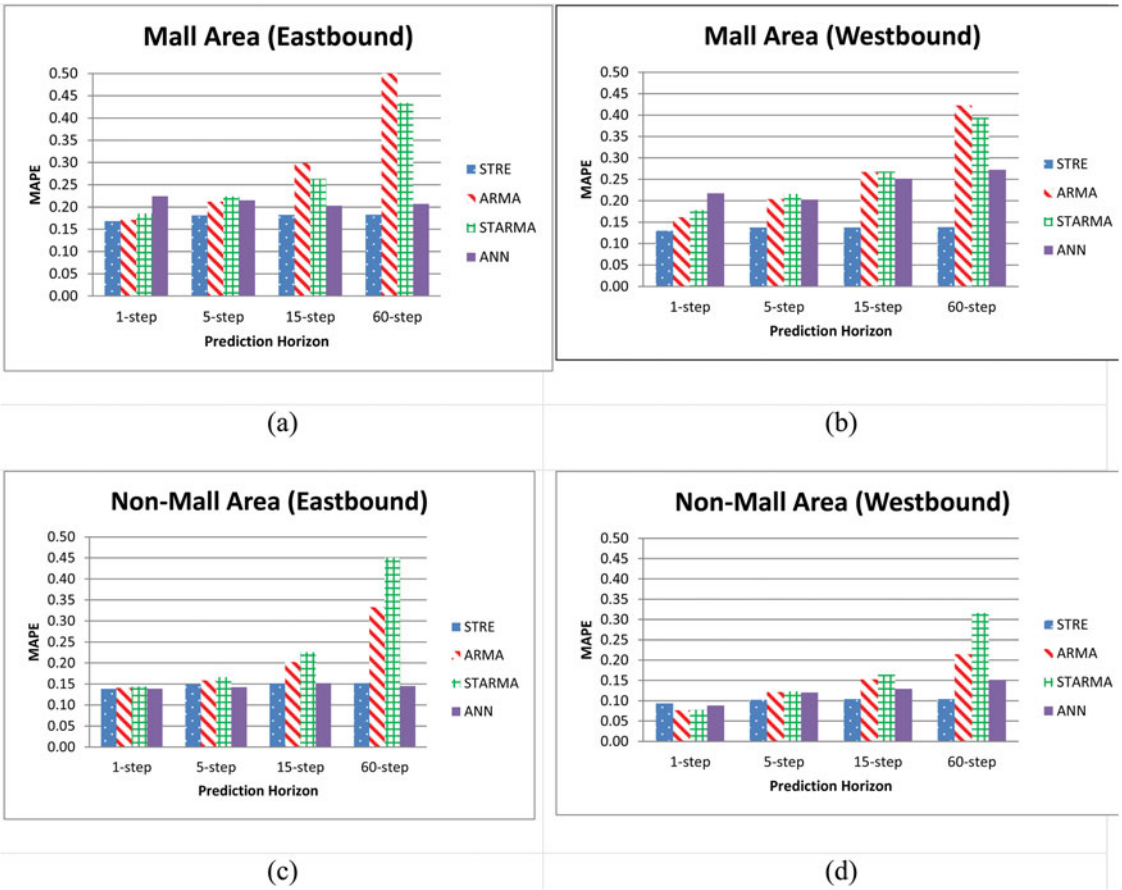

Figure 5 Comparisons of model prediction performance.

and STARMA models outperformed the traditional ones (Min, Hu, & Zhang, 2010). For ANN, a separate neural network model must be trained for each detector based on its individual traffic volume data time series, where the volume prediction for the next step is a nonlinear function of the previous five observations. The nonlinear function consists of various network parameters, including the number of hidden layers. These parameters are estimated by a standard 10-fold cross-validation function. The trained neural network can then be used to test the performance of the nonlinear prediction model for the traffic volume prediction problem considered in our study.

The MAPE results are compared in Figure 5, grouped by different traffic movement directions and areas. Each "bar" represents the average prediction MAPE for each prediction method. For the mall area, the prediction performance for STRE is stable for eastbound traffic because all the MAPEs are within 20% regardless of the prediction horizon, whereas the MAPEs of the other prediction methods increase significantly as the prediction horizon extends outward. As discussed earlier, the mall entrance significantly pollutes the data because the outbound and inbound traffic is fairly random at this location. For the mall area (westbound), STRE also demonstrates its robustness because the MAPE is within 13.4% regardless of the prediction horizon.

In terms of the non-mall area, STRE stably predicts traffic volume with low MAPEs (on average, 15% and 10% for eastbound and westbound, respectively, across all prediction horizons). This contrasts with the MAPEs for ARMA and STARMA, which increase significantly as the prediction horizon increases. For the non-mall area (eastbound), STARMA clearly experiences difficulties for the longer predictions; here, the MAPE reaches almost 45% for the 60-step prediction. This may be because the STARMA model as implemented does not utilize the volume data collected from the adjacent traffic detectors at Step 60 as the STRE model does for prediction, but instead considers only the temporal effect for the volume detector of interest and the spatial volume information collected in the past to perform its volume prediction. It is therefore reasonable that STRE outperforms STARMA in terms of long-term prediction.

Generally, the MAPEs for ANN grow as the prediction horizon is extended except for the non-mall area (eastbound) case, where ANN and STRE have equivalent performance (both average MAPEs are 14%) and statistically neither outperforms the other at a 5% significance level. ANN tends to predict a weighted average of previous volumes adjusted by a constant average value based on its historical observations, and this constant volume value is generally the average of historical traffic volume. The results indicate that ANN is capable of capturing

the nonlinear trend of the traffic volume data using a nonlinear prediction function trained based on historical data, but as the prediction horizon increases, ANN tends to predict a constant average value with small fluctuations. This results in large RMSEs but low MAPEs in any area where traffic does not fluctuate often such as the non-mall area, where traffic volume is not as volatile as the mall area and ANN can thus produce satisfactory results compared to STRE. However, where traffic is more unpredictable, such as around the mall, STRE is able to utilize adjacent detectors to estimate traffic volume even when the data are absent for as long as 60 steps, while ANN's prediction suffers due to its reliance on previous experience learning from the historical data.

Overall, these results demonstrate the robustness of the STRE model for both short-term and long-term prediction because the model is able to leverage the data output from adjacent sensors. Both ARMA and STARMA exhibit significant negative trends as the prediction horizon increases. For the 1-step case, ARMA (average MAPE = 13.5%) slightly outperforms STRE (average MAPE = 14.0%), although according to a paired *t*-test, this difference is not significant at a 5% significance level. The average MAPE for ANN (16.5%) is statistically greater than STRE's (14.0%) at a 5% significance level. For the 5-step, 15-step, and 60-step cases, the MAPEs of STRE are all statistically different from those of ARMA, STARMA and ANN at a 1% significance level. This evidence confirms our finding that the proposed STRE model outperforms ARMA, STARMA, and ANN.

CONCLUSIONS AND RECOMMENDATIONS

Predicting traffic on an urban traffic network using spatio-temporal models is a popular research area, and this article proposes an STRE model that can predict traffic volume by considering many detectors simultaneously. The city of Bellevue, in Washington state, was selected as the test site because it has installed more than 700 detectors that cover the entire city. Of these, 105 detectors are included in the modeling process described here, and the detectors on 8th Ave between a large shopping mall and freeway exit are used to demonstrate the prediction capability of the STRE model. This area is one of the busiest streets in the city and thus represents several challenging scenarios that can be used to test the proposed model. Our results show that the STRE model not only effectively predicts traffic volume but also outperforms three well-established volume prediction models, ARMA, STARMA, and ANN. Even without further model tuning, all the experimental links produced MAPEs between 8% and 16% except for those in three atypical locations, Link 45 (overall MAPE $\approx$ 29%), Link 75 (overall MAPE $\approx$ 21%), and Link 215 (overall MAPE $\approx$ 20%). As discussed earlier, the predictions for these locations could be potentially improved if the regional traffic patterns are considered in the basis function adjustment process. As shown in previous research (Tan, Wong, Xu, Guan, & Zhang, 2009), most other algorithms result in MAPEs ranging from 6% to 20%. Considering the highly volatile nature of our test network, small data aggregation level (5 min), active interactions between blocks, and long prediction horizons, these results are very encouraging. Compared with the ANN, ARMA, and STARMA models, the STRE model clearly demonstrates its prediction reliability and consistency over different prediction horizons. The ARMA and STARMA models, in particular, tend to fail when applied to long-horizon prediction. The results are consistent with finding from previous research.

Even though the STRE model provides encouraging prediction results, many challenges remain, the most important of which is the need to adjust many parameters during the calibration process. Preknowledge of traffic patterns would facilitate the model-tuning process. Future model improvement could take many potential directions. First, investigating how best to decide the number of reference points and locations is an issue worth further study. The relationship between the accuracy and the number of the reference points is worth investigation. Second, the selection of the basis function is critical, and once this has been determined, the subsequent tune-up process will also present a challenge for researchers. One possible approach might be to change the proposed penalty function in the basis function. A case-by-case basis might also significantly improve the results, especially for Links 215 and 45, both of which underperformed in the study. Through-movement detectors alone are used in this study, but if turning-movement counts are also incorporated, the penalty value can be more precisely determined to increase the prediction accuracy.

ACKNOWLEDGMENTS

The authors thank Xinyu Min for providing his source code to complete the comparisons. We are also grateful for data support from the Smart Transportation and Application Research Laboratory (STAR Lab) at the University of Washington and to the city of Bellevue, WA.

REFERENCES

Chandra, S. R., & Al-Deek, H. (2009). Predictions of freeway traffic speeds and volumes using Vector Autoregressive Models. *Journal of Intelligent Transportation Systems*, **13**(2), 53–72.

Cheng, T., Haworth, J., & Wang, J. (2011). Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems*, **14**(4), 389–413.

Chiou, Y.-C., Lan, L. W., & Tseng, C.-M. (2014). A novel method to predict traffic features based on rolling self-structured traffic patterns. *Journal of Intelligent Transportation Systems*, **18**(4), 352–366.

Cressie, N., Shi, T., & Kang, E. L. (2010). Fixed rank filtering for spatio-temporal data. *Journal of Computational and Graphical Statistics*, **19**(3), 724–745.

Cressie, N., & Wikle, C. K. (2002). Space–time Kalman filter. *Encyclopedia of Environmetrics*, **4**, 2045–2049.

Hamed, M. M., & Al-Masaeid, H. R. (1995). Short-term prediction of traffic volume in urban arterials. *Journal of Transportation Engineering*, **121**(3), 249–254.

Ghosh, B., Basu, B., & O'Mahony, M. (2009). Multivariate short-term traffic flow forecasting using time-series analysis. *IEEE Transactions on Intelligent Transportation Systems*, **10**(2), 246–254.

Kamarianakis, Y., Kanas, A., & Prastacos, P. (2005). Modeling traffic volatility dynamics in an urban network. *Transportation Research Record*, **1923**(1), 18–27.

Kamarianakis, Y., & Prastacos, P. (2003). Forecasting traffic flow conditions in an urban network comparison of multivariate and univariate approaches. *Transportation Research Record: Journal of the Transportation Research Board*, **1857**(2003), 74–84.

Katzfuss, M., & Cressie, N. (2011). Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, **32**(4), 430–446.

Ma, X., Wu, Y.-J., & Wang, Y. (2011). DRIVE net: An E-science of transportation platform for data sharing, visualization, modeling, and analysis. *Transportation Research Record*, **2215**, 37–49.

Math Works. (2013). *Neural network toolbox*. Retrieved from http://www.mathworks.com/products/neural-network/

Min, W., & Wynter, L. (2011). Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, **19**(4), 606–616.

Min, X., Hu, J., Chen, Q., Zhang, T., & Zhang, Y. (2009). Short-term traffic flow forecasting of urban network based on dynamic STARIMA model. *12th International IEEE Conference on Intelligent Transportation Systems, 2009, ITSC '09*, 1–6.

Min, X., Hu, J., & Zhang, Z. (2010). Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model. *13th International IEEE Conference on Intelligent Transportation Systems*, 1535–1540.

Qiao, W., Haghani, A., & Hamedi, M. (2013). A nonparametric model for short-term travel time prediction using bluetooth data. *Journal of Intelligent Transportation Systems*, **17**(2), 165–175.

Smith, B. L., & Demetsky, M. J. (1997). Traffic flow forecasting: Comparison of modeling approaches. *Journal of Transportation Engineering*, **123**(4), 261–266.

Smith, B. L., Williams, B. M., & Keith Oswald, R. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, **10**(4), 303–321.

Stathopoulos, A., & Karlaftis, M. G. (2003). A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, **11**(2), 121–135.

Tan, M.-C., Wong, S., Xu, J.-M., & Guan, Z.-R. (2009). An aggregation approach to short-term traffic flow prediction. *IEEE transactions on Intelligent Transportation Systems*, **10**(1), 60–69.

Thomas, T., Weijermars, W., & van Berkum, E. (2010). Predictions of urban volumes in single time series. *IEEE Transactions on Intelligent Transportation Systems*, **11**(1), 71–80.

Vlahogianni, E. (2009). Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics. *Journal of Intelligent Transportation Systems*, **13**(2), 73–84.

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2007). Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks. *Computer-Aided Civil and Infrastructure Engineering*, **22**(5), 317–325.

Washington, S., Karlaftis, M. G., & Mannering, F. L. (2003). *Statistical and econometric methods for transportation data analysis*. Boca Raton, FL: Chapman & Hall/CRC.

Williams, B. M., & Hoel, L. A. (2003). Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results. *Journal of Transportation Engineering*, **129**(6), 664–672.

Wu, Y.-J., An, S., Ma, X., & Wang, Y. (2011). Development of web-based analysis system for real-time decision support on arterial networks. *Transportation Research Record*, **2215**, 24–36.

Xie, Y., Zhang, Y., & Ye, Z. (2007). Short-term traffic volume forecasting using Kalman Filter with discrete wavelet decomposition. *Computer-Aided Civil and Infrastructure Engineering*, **22**(5), 326–334.

Zou, H., Yue, Y., Li, Q., & Shi, Y. (2010). A spatial analysis approach for describing spatial pattern of urban traffic state. *13th International IEEE Conference on Intelligent Transportation Systems*, 557–562.