Traffic Flow Prediction for Urban Network using Spatio-Temporal Random Effects Model

Yao-Jan Wu (Corresponding Author)
Assistant Professor
Department of Civil Engineering
Parks College of Engineering, Aviation and Technology
3450 Lindell Boulevard, McDonnell Douglas Hall Room 2051
Saint Louis University
St. Louis, MO 63103
Office: 314-977-8249; Fax: (314)977-8388;
Email: yaojan@slu.edu

Feng Chen
Ph.D. Candidate
Department of Computer Science
Northern Virginia Graduate Center
Virginia Tech
7054 Haycock Road, Room 310
Falls Church, VA 22043

Chang-Tien Lu
Associate Professor
Department of Computer Science
Northern Virginia Graduate Center
Virginia Tech
7054 Haycock Road, Room 310
Falls Church, VA 22043
Tel: (703) 538-8373, Fax: (703) 538-8348

Brian Smith
Professor
Department of Civil and Environmental Engineering
University of Virginia
Thornton Hall B228
351 McCormick Road
Charlottesville, VA 22904-1000
Phone: (434) 243-8585
Email: briansmith@virginia.edu

Word count: 5547 + 4Figures and 1 Tables (1250 words) = 6797 words

Submitted for Presentation at the 91st Annual Meeting of the Transportation Research Board
Submitted on August 1, 2011

ABSTRACT

Traffic prediction is critical to success of Intelligent Transportation Systems (ITS). Predicting traffic on an urban traffic network using spatio-temporal models has become a popular research area in the past decade. The model does not only rely on observation data at the detector of interest but also takes advantage of neighboring detectors to provide better prediction capability. However, most models suffer high mathematical complexity and low flexibility in tune-up. This paper presents a novel Spatio-Temporal Random Effects (STRE) model that has a reduced computational complexity due to mathematical dimension reduction, and additional tune-up flexibility provided by the basis function that is able to take traffic patterns into account. The City of Bellevue, WA is selected as the model test site due to the widespread locations of the loop detector in the City. Data collected from 105 detectors in the downtown area during the first two weeks of July, 2007 are used in the modeling process and the traffic volumes are predicted for 14 detectors during the first week of July, 2008. The results not only show that the model can effectively consider the neighboring detectors to accurately predict the traffic in locations with regular traffic patterns, but also verify its temporal transferability. Except three special locations, all experimental links have Mean Average Percentage Errors (MAPEs) between 8% and 15%. Without further model tune-up, the results are encouraging.

1. INTRODUCTION

Reliable, accurate and consistent real-time traffic information is a key to success in the development and implementation of the Intelligent Transportation Systems (ITS). For example, the Advance Traveler Information System (ATIS), a subsystem of ITS, relies heavily on high quality real-time traffic data to provide road users with up-to-date guidance. Moreover, the Advance Traffic Management System (ATMS), another subsystem of ITS, also requires accurate traffic information to implement the traffic control schemes. In the past, the collection of real-time data was the foremost goal. Currently, most agencies have begun to consider taking advantage of the vast archived datasets for “real-time forward-looking analysis” *(1)*. With predicted data, proactive transportation management is feasible.

In today’s ITS environment, time- and location-specific data are collected in huge volumes in real-time *(2)* and more and more agencies are capable of archiving real-time traffic data. Processing real-time and historical traffic data simultaneously could provide useful results. Reliable short-term traffic prediction algorithms can provide many benefits to traffic management without further investment in new facilities. Unfortunately, a consistent data feed to the Traffic Management Center (TMC) is not always feasible. Inconsistent data connection is one of the key problems for arterial ATIS due to communication errors and malfunctioning detectors *(3)*. Maintaining consistent, high quality traffic data flow has been a challenging task for researchers and practitioners. A robust short-term traffic prediction is a key to successful ITS application.

Besides, travel demand forecasting also relies on short-term traffic flow prediction *(4)*. Over the past three decades, most efforts have focused on freeway traffic status (volume, speed or occupancy) prediction. For example, the work done by *(4,5,6)* demonstrates great efforts in traffic volume prediction. Many previous efforts are also summarized in these research papers. The urban networks are usually more complicated than freeways. Thus, there is greater likelihood of communication disruption. Moreover, the traffic control strategies would be less responsive because of the lag between traffic data detection and implementation. Due to the complex infrastructure of urban cities, a more responsive volume prediction scheme is required but is also more challenging. In terms of volume prediction method development, there are two major differences between freeways and arterials. First, the spatial locations of detectors are usually closer in arterials. The traffic prediction method being developed can take advantage of the geospatial relationships between detectors to provide better prediction accuracy. Second, the urban traffic suffers from delays caused by signalized intersections. Traffic status would introduce more irregularity and uncertainty to the traffic prediction because of different traffic characteristics, such as frequent occurrences of queues and lane-changing behaviors. These factors may lead to a low prediction precision on arterial networks. The traffic prediction method needs to be more responsive to react to the rapid changes in urban traffic status.

2. LITERATURE REVIEW

Despite the fact that the spatial relationships are strong and noticeable on urban networks, most research has focused on “one point” (or one detector) short-term traffic prediction in the past decades. In other words, the dependencies between detectors (spatial domain) were not considered but only the temporal domain was considered. These are also called “univarite” methods, which are similar to those used for freeway cases. Among all univariate methods, time series-based methods are considered most popular. The autoregressive integrated moving average (ARIMA)-based method is commonly used, e.g. *(6)* and *(7)*. Many of the univariate models are compared by many researchers. For example, Smith and Demetsky *(5)* compared historical average, time series (ARIMA), back-propagation neural network, and nonparametric regression. Later on, Smith et al. *(8)* found the results generated by seasonal ARIMA models are statistically superior to those produced by non-parametric regression. In general, the ARIMA-based models yield satisfactory performance.

Most recently, univariate time series-based approaches are still used to predict traffic but with improvement. For example, *(9)* 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 naïve ARIMA, nonparametric regression and NN models. Thomas et al. *(10)* developed a heuristic approach to predict short- and long-term traffic. The novelty of their method relies on the mixed method combining the concept of time-series (temporal correlation) and the application of Kalman filter (to reduce the noise).

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

In the past decade, more and more research efforts further considered spatial information to improve prediction accuracy. Among all the methods, multivariate time series have been popular, such as Spatio-Temporal (ST) ARIMA *(11,12,13)*, multivariate structural time-series *(14),* Dynamic STARIMA *(15)* and Generalized STARIMA *(16).* However, time series models have many parameters to calibrate. Smith et al. (4) compared several parametric and nonparametric traffic prediction models and found the ARIMA model is fairly time consuming. Due to the nature of multi-variate time series, adding one more dimension (spatial) would increase computational complexity and estimation of a large number of parameters *(14).*

Recently, the spatio-temporal correlations have gained more attention and been used to forecast traffic flow. Vlahogianni et al. (17) developed a traffic volume predictor that uses "temporal structures of feed-forward multilayer perceptrons (MLP)." Later on, Vlahogianni *(18)* further enhance the pattern-based neural network prediction scheme by considering traffic flow regimes. Zou et al. *(19)* use a spatial autocorrelation method to estimate the patterns of traffic states among urban streets based on historical travel time data. However, traffic flow prediction is not a focus in this research. Cheng et al. *(2)* investigate 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 developed in *(1)* is designed to minimize the number of parameters, reducing the computational costs. This allows the model to be applied to large metropolitan areas. The model was tested on a large urban network.

Based on the literature review, the common challenge for the spatio-temporal model-related research is the dimension of the network. Once the network grows, most spatio-temporal models are not capable of handling a large network in a timely manner. Huge datasets collected from a large network become more and more common with the rapid development and implantation of ITS sensors. A large number of spatial detectors would result in a highdimensional statistical model. To deal with this issue, a Spatio-Temporal Random Effects (STRE) model is adopted in this study to handle these issues.

3. METHODOLOGY

Inheriting the filtering capability of Kalman Filter, the Spatio-Temporal Kalman Filter (STKF) expands KF to a spatio-temporal domain. However, the traditional STKF suffers from its low performance in modeling high-dimension data *(20).* The STRE model, a special type of STKF, is proposed by Cressie et al. *(20)* and has proven its mathematical effectiveness in dimension rededuction and parameter estimation *(20).* To explain this 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, e.g. parameter estimation and prediction process, will be also elaborated.

3.1 Spatial Random Effects (SRE) Model

Let $(\mathrm{Y}(\mathbf{s})\colon \mathrm{s} \in \mathrm{D} \in \mathbb{R}^2)$ be a real-valued spatial process. The Spatial Random Effects (SRE) model first decomposes the spatial process into two additive components

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

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

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

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

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

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

where $\mathbf{b}(\mathbf{s}) = \left[\mathrm{b}_1(\mathbf{s}), \ldots, \mathrm{b}_\mathrm{p}(\mathbf{s})\right]^{\mathbf{T}}$**,** $\boldsymbol{\eta}$ is a vector of r-dimensional Gaussion process with mean zero and co-varians **K:** $\boldsymbol{\eta} \sim \boldsymbol{\mathcal{N}}_\mathrm{r}(\mathbf{0}, \mathbf{K})$**,** and $\xi(\mathbf{s})$ is independent Gaussian white noise with zero mean and variance σ^2_ξ. Then, by combining Equations (3.1), (3.2), and (3.3), we have the SRE Model as

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

The unknown parameters are $\{\boldsymbol{\beta}, \sigma^2_{\epsilon,t}, \sigma^2_\xi\}$. It is shown that by employing this form, the resulting Best Linear Unbiased Predictor (BLUP) could achieve significant computational savings compared with a traditional Kriging Model. The BLUP estimator based on the SRE model is also named fixed-rank Kriging.

3.2 Spatio-Temporal Random Effects Model (STRE)

The Spatio-Temporal Random Effects (STRE) model is regarded as the extension of the SRE model with consideration of temporal effects. The STRE model can perform the following tasks: dimension reduction (spatial) and rapid smoothing, filtering, or prediction (temporal) *(20)*. The filtering, smoothing, and prediction based on STRE are also named fixed rank filtering (FRF), fixed rank smoothing, and fixed rank prediction *(20)*.

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

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

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

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

$$\mathbf{Z}_t = \mathbf{O}_t \mathbf{Y}_t + \boldsymbol{\epsilon}_t, t = 1,2, \dots, \tag{6}$$

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

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

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

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

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

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

where $\mathbf{B}_t = \left[\mathbf{b}_t(\mathbf{s}_{1,t}), \dots, \mathbf{b}_t(\mathbf{s}_{n_t,t})\right]^{\mathrm{T}}$, $\mathbf{b}_t(\mathbf{s}_{1,t}) = \left[\mathrm{b}_{1,t}(\mathbf{s}_{1,t}), \dots, \mathrm{b}_{r,t}(\mathbf{s}_{1,t})\right]$ 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 $\mathrm{r} \times \mathrm{r}$ covaraince matrix given by $\mathbf{K}_t$. The first component in (8) denotes a smoothed small-scale variation at time t, captured by the set of basis functions $\{\mathbf{b}_t(\mathbf{s}_{1,t}), \ldots, \mathbf{b}_t(\mathbf{s}_{n_t,t})\}$.

The second component in (8) captures the fine-scale variability similar to the nugget effect as defined in geostatistics (*20*). It is assumed that $\boldsymbol{\xi}_t \sim \boldsymbol{\mathcal{N}}_{m_t}(0, \sigma^2_{\xi,t}\mathbf{V}_{\xi,t})$, $\mathbf{V}_{\xi,t} = \operatorname{diag}\left(v_{\xi,t}(\mathbf{s}_{1,t}), \ldots, v_{\xi,t}(\mathbf{s}_{n_t,t})\right)$, and $v_{\xi,t}(\cdot)$ 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 $v_{\xi,t}(\cdot) = \exp\{\mathbf{b}_\xi^T \boldsymbol{\eta}_\xi\}$, where $\mathbf{b}_\xi$ is a vector of r_ξ basis functions, with $r_\xi < r$ (*21*). 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 one,

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

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

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

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

where $\boldsymbol{\mu}_t$ is deterministic and the other components are stochastic (*20*).

3.3 Filtering, Smoothing and Prediction

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

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

$$\mathbf{Y}_{t|t} = \mathbf{O}_t\boldsymbol{\mu}_t + \mathbf{O}_t\mathbf{B}_t\boldsymbol{\eta}_{t|t} + \mathbf{O}_t\boldsymbol{\xi}_{t|t}, \tag{11}$$

$$\boldsymbol{\eta}_{t|t} = \boldsymbol{\eta}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{B}_t^T\mathbf{O}_t^T\left[\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1}\mathbf{B}_t^T\mathbf{O}_t^T + \mathbf{D}_t\right]^{-1}\left(\mathbf{z}_t - \mathbf{O}_t\mathbf{X}_t\boldsymbol{\beta}_t - \mathbf{O}_t\mathbf{B}_t\boldsymbol{\eta}_{t|t-1}\right),$$

$$\boldsymbol{\xi}_{t|t} = \sigma^2_{\xi,t}\mathbf{V}_{\xi,t}\mathbf{O}_t^T\left[\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1}\mathbf{B}_t^T\mathbf{O}_t^T + \mathbf{D}_t\right]^{-1}\left(\mathbf{z}_t - \mathbf{O}_t\mathbf{X}_t\boldsymbol{\beta}_t - \mathbf{O}_t\mathbf{B}_t\boldsymbol{\eta}_{t|t-1}\right),$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{B}_t^T\mathbf{O}_t^T\left[\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1}\mathbf{B}_t^T\mathbf{O}_t^T + \mathbf{D}_t\right]^{-1}\mathbf{O}_t\mathbf{B}_t\mathbf{P}_{t|t-1},$$

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

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

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

$$\mathbf{Y}_{t|T} = \mathbf{O}_t\boldsymbol{\mu}_t + \mathbf{O}_t\mathbf{B}_t\boldsymbol{\eta}_{t|T} + \mathbf{O}_t\boldsymbol{\xi}_{t|T} \tag{12}$$

$$\boldsymbol{\eta}_{t|T} = \boldsymbol{\eta}_{t|t} + \mathbf{J}_t\left(\boldsymbol{\eta}_{t+1|T} - \boldsymbol{\eta}_{t+1|t}\right)$$

$$\xi_{t|T} = \xi_{t|t} - \mathrm{M}_t(\eta_{t+1|T} - \eta_{t+1|t})$$
$$\mathrm{P}_{t|T} = \mathrm{P}_{t|t} + \mathrm{J}_t(\mathrm{P}_{t+1|T} - \mathrm{P}_{t+1|t})\mathrm{J}_t^{\mathrm{T}}$$
$$\mathrm{R}_{t|T} = \mathrm{R}_{t|t} + \mathrm{M}_t(\mathrm{P}_{t+1|T} - \mathrm{P}_{t+1|t})\mathrm{M}_t^{\mathrm{T}},$$

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

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

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

Computational Complexity

The computational complexity is calculated based the total number of observed time stamps, the total number of observed spatial locations $\mathrm{n_t}$ at time t, and the number of bases used in the hidden process $\{\boldsymbol{\eta}_{\mathrm{t}}\}$. We compare the computational complexity between the traditional spatio-temporal Kalman filtering (STKF) model (*22*) and the STRE model. Given observed data $\{\mathbf{z}_1, \ldots, \mathbf{z}_{\mathrm{t}}\}$, the computational complexity of the spatio-temporal Kalman filtering is $\mathrm{O}(\sum_{\mathrm{t}}\mathrm{n_t^3})$. In comparison, the computational complexity of the fixed-rank filtering based on the STRE model is $\mathrm{O}(\sum_{\mathrm{t}}\mathrm{n_t}\mathrm{r}^3)$. In general, r is fixed with $\mathrm{r} \ll \mathrm{n}$. Therefore, we have the computational complexity for the STRE model as $\mathrm{O}(\sum_{\mathrm{t}}\mathrm{n_t})$, which is linear order complexity. The comparison results indicate STRE model achieves significant computational savings, compared with traditional STKF.

3.4 Parameter Estimation

For the model parameter estimation process, the Expectation-Maximization (EM) algorithm (*21*) is used. We first assume that the parameter $\sigma_{\epsilon,t}^2$ is known, and focus on the estimation of the parameters $\boldsymbol{\theta} = \{\beta_{\mathrm{t}}, \sigma_{\xi,\mathrm{t}}^2, \mathrm{H_t}, \mathrm{U_t}, \mathrm{K_0}\}$. The STRE model only depends on the parameters $\mathrm{H_t}$ and $\mathrm{U_t}$ through the relationships $\eta_{\mathrm{t|t-1}} = \mathrm{H_t}\eta_{\mathrm{t-1|t-1}}$ and $\mathrm{P_{t|t-1}} = \mathrm{H_t}\mathrm{P_{t-1|t-1}}\mathrm{H_t^T} + \mathrm{U_t}$. It implies that there will be no unique MLE estimation if both $\mathrm{H_t}$ and $\mathrm{U_t}$ are allowed to be different at different time stamps. Hence, to achieve the identifiability of the parameters, it is assumed that $\mathrm{H} = \mathrm{H_1} = \cdots = \mathrm{H_T}$ and $\mathrm{U} = \mathrm{U_1} = \cdots = \mathrm{U_T}$. The complete negative log likelihood function is

$$-2\log \mathrm{L_c}(\boldsymbol{\theta}) = -2\log \mathrm{f}(\mathbf{z}_{1:\mathrm{T}}, \boldsymbol{\eta}_{1:\mathrm{T}}, \boldsymbol{\xi}_{1:\mathrm{T}}|\boldsymbol{\theta})$$
$$= \sum_{\mathrm{t=1}}^{\mathrm{T}}\mathrm{tr}\left(\mathbf{V}_{\epsilon,t}^{-1}[\mathbf{z}_t - \mathbf{X}_t\boldsymbol{\beta}_t - \mathbf{B}_t\boldsymbol{\eta}_t - \boldsymbol{\xi}_t][\mathbf{z}_t - \mathbf{X}_t\boldsymbol{\beta}_t - \mathbf{B}_t\boldsymbol{\eta}_t - \boldsymbol{\xi}_t]^{\mathrm{T}}\right)/\sigma_{\epsilon,t}^2$$

$$+\sum_{t=1}^{T} n_t \log \sigma_{\xi,t}^2 + \frac{\sum_{t=1}^{T} \mathrm{tr}\left(\mathbf{V}_{\xi,t}^{-1} \boldsymbol{\xi}_t \boldsymbol{\xi}_t^{T}\right)}{\sigma_{\xi,t}^2} + \log|\mathbf{K}_0| + \mathrm{tr}(\mathbf{K}_0^{-1}\, \boldsymbol{\eta}_0 \boldsymbol{\eta}_0^{T}) + T\log|\mathbf{U}|$$
$$+\sum_{t=1}^{T} \mathrm{tr}(\mathbf{U}^{-1}[\boldsymbol{\eta}_t - \mathbf{H}\boldsymbol{\eta}_{t-1}][\boldsymbol{\eta}_t - \mathbf{H}\boldsymbol{\eta}_{t-1}]^{T}) + \mathrm{const}$$

Let $\mathbf{K}_t^{[l+1]} = \mathbf{P}_{t+T}^{[l]} + \boldsymbol{\eta}_{t|T}^{[l]} \boldsymbol{\eta}_{t|T}^{[l]\,T}$ and $\mathbf{L}_t^{[l+1]} = \mathbf{P}_{t,t-1|T}^{[l]} + \boldsymbol{\eta}_{t|T}^{[l]} \boldsymbol{\eta}_{t|T}^{[l]\,T}$. The EM updates of the parameters are as follows:

$$\boldsymbol{\beta}_t^{[l+1]} = \left(\mathbf{X}_t^{T} \mathbf{V}_{\epsilon,t}^{-1} \mathbf{X}_t\right)^{-1} \mathbf{X}_t^{T} \mathbf{V}_{\epsilon,t}^{-1} \left[\mathbf{z}_t - \mathbf{B}_t \boldsymbol{\eta}_{t|T}^{[l]} - \boldsymbol{\xi}_{t|T}^{[l]}\right]$$
$$\sigma_{\xi,t}^2 = \mathrm{tr}\left(\mathbf{V}_{\xi}^{-1}\left[\boldsymbol{R}_{t|T}^{[l]} + \boldsymbol{\xi}_{t|T}^{[l]} \boldsymbol{\xi}_{t|T}^{[l]\,T}\right]\right)/n_t \tag{14}$$
$$\mathbf{K}_0^{[l+1]} = \mathbf{P}_{0|T}^{[l]} + \boldsymbol{\eta}_{0|T}^{[l]} \boldsymbol{\eta}_{0|T}^{[l]\,T}$$
$$\mathbf{H}^{[l+1]} = \left(\sum_{t=1}^{T} \mathbf{L}_t^{[l+1]}\right)\left(\sum_{t=0}^{T-1} \mathbf{K}_t^{[l+1]}\right)^{-1}$$
$$\mathbf{U}^{[l+1]} = \frac{\left(\sum_{t=1}^{T} \mathbf{K}_t^{[l+1]} - \mathbf{H}^{[l+1]} \sum_{t=1}^{T} \mathbf{L}_t^{[l+1]\,T}\right)}{T}.$$

The EM algorithm for the STRE model is as follows:

Step 1: Select an initial estimate of the parameters $\boldsymbol{\theta}^{[0]}$

Step 2: For $l = 1, \ldots,$ until convergence

Step 2.1 Use the parameters $\boldsymbol{\theta}^{[l]}$ smoothing estimators in Equation (12) to estimate $\boldsymbol{\eta}_{t|T}^{[l]}, \boldsymbol{\xi}_{t|T}^{[l]}, \boldsymbol{R}_{t|T}^{[l]}$, and $\mathbf{P}_{t|T}^{[l]}$.

Step 2.2 Use Equation (13) to obtain the updated $\boldsymbol{\theta}^{[l+1]}$

4. STUDY SITE AND DATA COLLECTION

The traffic volume data are collected in the City of Bellevue, Washington (WA). The traffic volume data are collected from the advance loop detector, which is located 100 ~ 130 feet (30.5 ~ 39.7 m) upstream from the stop bar at each approach. As of July 2010, the City has more than 182 signalized intersections, 165 of which are controlled by 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 (*3, 23*) at the Smart Transportation Application and Research Laboratory (STAR Lab) at the University of Washington (UW), Seattle.

This study focuses 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 take advantage of the high correlation between detectors due to proximity of intersections. The downtown area in Figure 1 is selected as our test site. In total, 105 detectors in this area are included in the modeling process. 8th Ave is selected as the test route because it is a fairly busy street, with annual average weekday traffic of 37,700 (veh/day), connecting freeway I-405 and a large shopping mall (Bellevue Square). 14 detectors, seven eastbound and seven westbound, on 8th Ave are used to examine the model's capability. Since each link only has only one detector, a link also represents a detector hereafter in this study. These links and reference points used in this study are illustrated in Figure 1. The reference point is overlapped with the intersection number and its concept will be explained in the next section.

Weekday data (Tuesday, Wednesday and Thursday) collected from first two weeks of June, 2007 are used for training and the last two weeks of June, 2007 are used for cross validation. The verification data are collected during the first week of July in 2008. In this study, all data are aggregated into 5-minute intervals to reduce the effect of random noise.

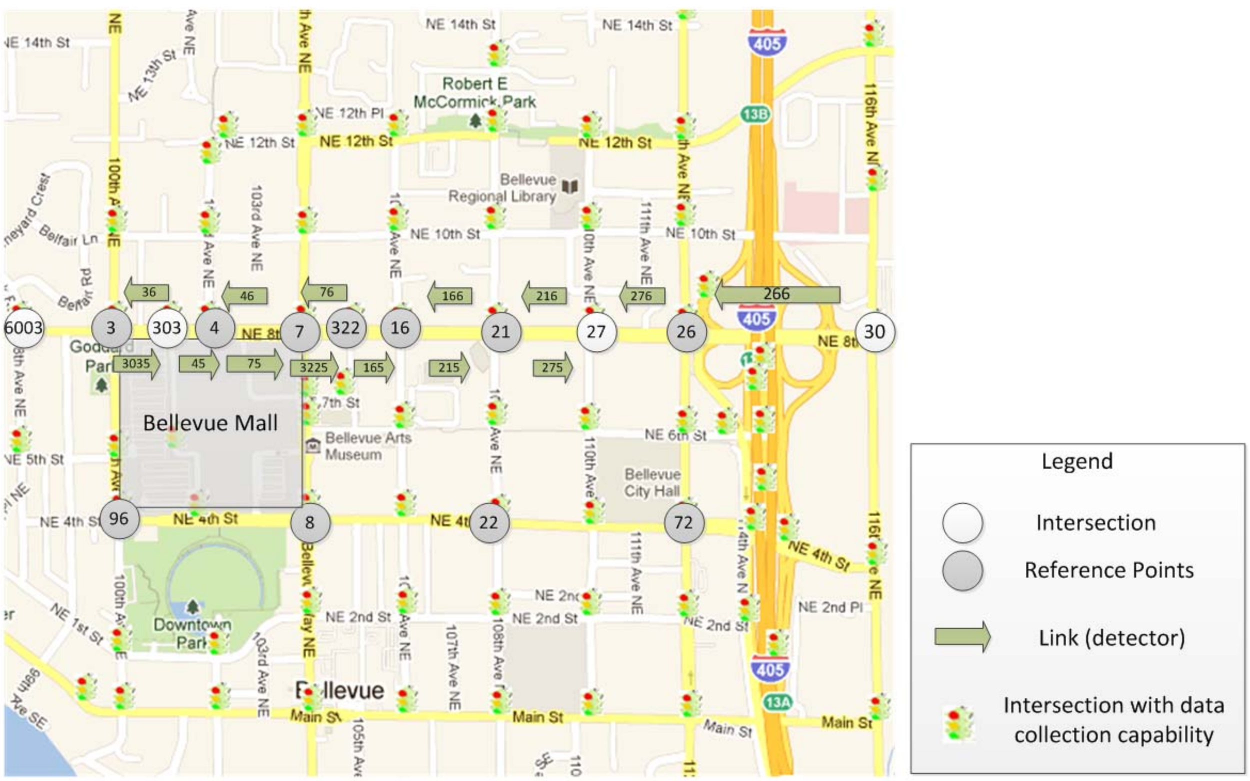


Figure 1 Downtown area in the City of Bellevue, WA (background images are from maps.google.com)

5. MODEL ADJUSTMENT

Before the modeling process, the basis functions in Equation (8) have to be determined. As to the selection of basis function, the bisquare function is used in this study and is defined as: .

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

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

The range parameter determines the independency between two links. The smaller the range parameter is, the more likely two links are independent. Based on the experimental results, the east-west distance between downtown boundaries is considered most suitable for our study.

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 reference points are critical to prediction accuracy. The reference point is set as the point between two detectors. Since the detectors are assumed to be in the middle of the link, each reference point is located at the intersection (node) in this study. As found in our experiment, the more reference points are included in the analysis, the better the results will be. However, computational efficiency will decrease. In order to increase model performance, the number of center points needs to be relatively small. In the experiments, 11 reference points are considered and illustrated in Figure 1 in a dark color. It should be noted that, different from regular spatio-temporal data, the data collected in a transportation network need to consider the direction of traffic flow. Two links with reversed directions between same pair of intersections would overlap with each other. Therefore, the reference points determined by these two pairs of links will be also overlap (at the same intersection), but with opposite directions.

Directional Penalty

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

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

Note that the greater the penalty value (basis function), the lower the correlation between the reference point and the detector. In this case, the detector would contribute less volume to the reference point. Take the intersection in Figure 2 for example. To determine the penalty for the contribution of a detector to the reference point with 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; i.e. 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; i.e. 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; i.e. 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; i.e. Detector 4 or Detector 5 has minor contribution to the reference point. This is because the traffic detected on Detector 1 is less likely to be collected by Detector 4 or Detector 5 again since only through traffic detectors are used in this study.

Note that all the penalties are adjusted based on the results from the cross validation.

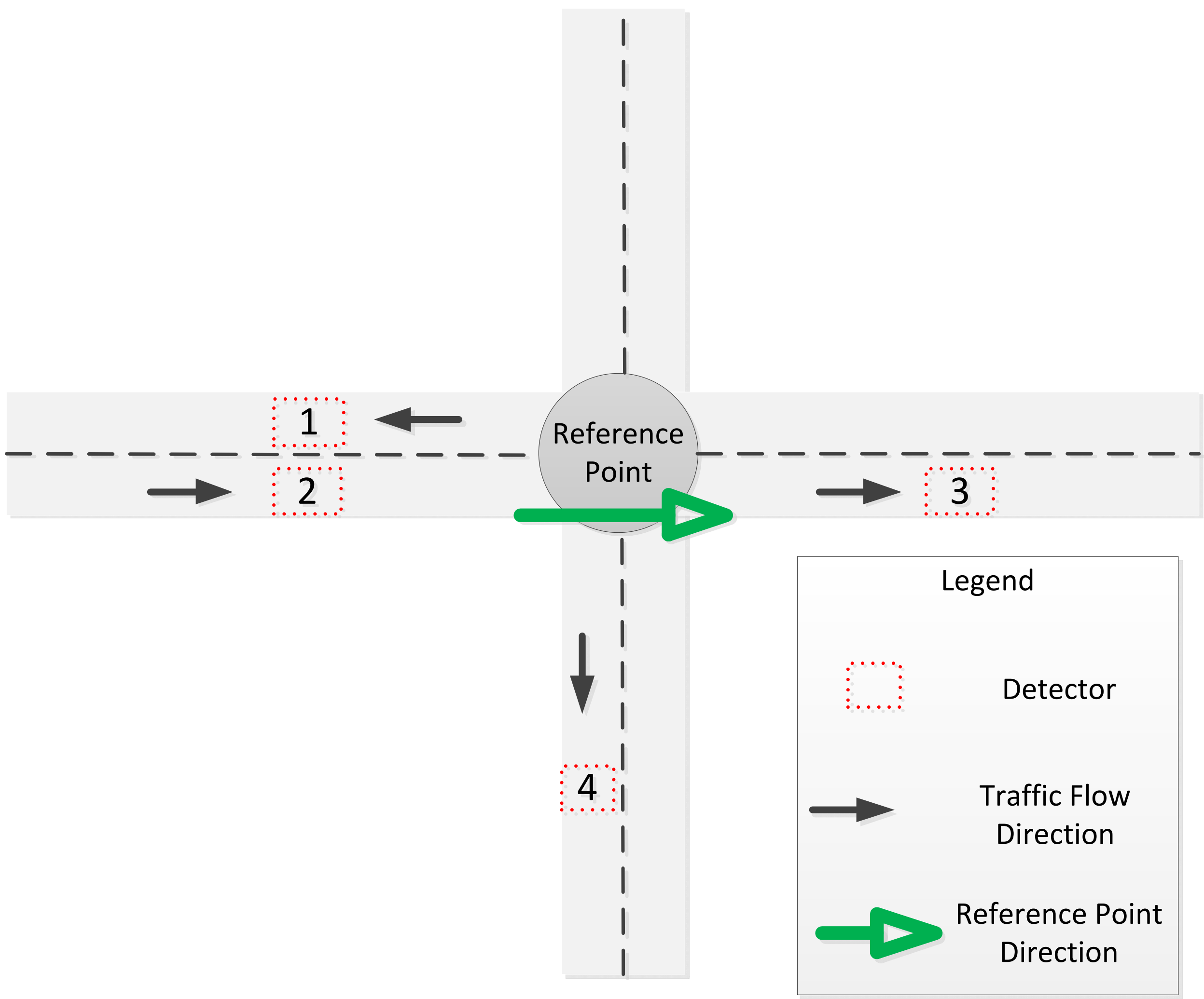


FIGURE 2 Basis function penalty assignment

6. PREDICTION PERFORMANCE

6.1 Performance Indexes

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 *(9, 25, 27)* and are defined as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} \left| \frac{G_t - Z_t}{G_t} \right| \tag{15}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(G_t - Z_t)^2}{n}} \tag{16}$$

6.2 Experimental Results

To evaluate the temporal transferability of the STRE model, the model was verified with the data collected during the first week of July, 2008 (Tuesday, Wednesday and Friday). In our analysis, the prediction results of all the links are separated into groups: mall area and non-mall area because these two areas have different traffic patterns. The reference point, 322, is regarded as a separation boundary to separate two kinds of traffic patterns. The mall area (Bellevue Square) has about 180 retail stores and more than 10,000 parking spaces. This shopping mall attracts more than 43,000 visitors daily. Therefore, the parking lots around the mall create irregular traffic patterns that might disturb the spatio-temporal prediction accuracy.

Table 1 shows the model verification results divided by two areas. Scenarios of 1, 5, 15 and 60 step prediction horizons are adopted. As expected, the prediction accuracy degrades as the prediction horizon increases. However, the prediction accuracy only degrades slightly. Overall, the prediction results are satisfying. Figures 3(a) and 3(b) show the example results of Links 165 and 36, respectively. These two figures show two distinct patterns in the downtown area and demonstrate the challenges in our datasets. This results shows the STRE model is adaptive to many traffic patterns.

In terms of prediction accuracy for different areas, the prediction MAPEs in the non-mall area are between 11.6% (one-step) and 12.5% (60-step) while the MAPEs in the mall area are between 16.9% and 17.5%. The resulting RMSEs follow the same trend of MAPEs. In the non-mall area, the overall prediction accuracy is satisfying (most MAPEs ≈11 %). However, the STRE model tends to overestimate the volume on Link 215 (MAPE≈20%), as shown in Figure 4(a). This link was a special data collection point where the City estimates the volume by its upstream and downstream detectors. In other words, the ground truth data being used are still estimated values. For the links in the mall area, the prediction of Link 45 has the lowest performance. The result is not surprising because the link is located at the major entrance and exit of the parking lot. The traffic pattern there is fairly unstable. As shown in Figure 4(b), the

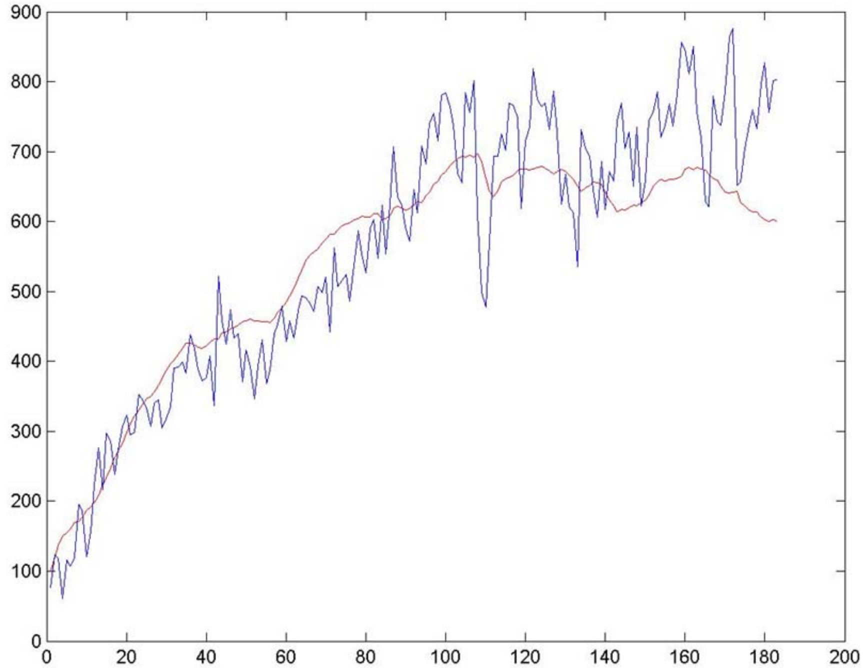
STRE model tends to underestimate the volume because the model might be unable to capture the volume from the parking lot.

For both mall or non-mall areas, the westbound traffic prediction is consistently better than the east bound one. It is very likely that the traffic control coordination system is designed to favor westbound traffic. The semi-actuated coordinated signal control scheme is implemented on 8^{th} Ave to release traffic from the off ramp on the freeway (I-405). This finding suggests the consideration of traffic control scheme should be considered in the future modeling tuning process.

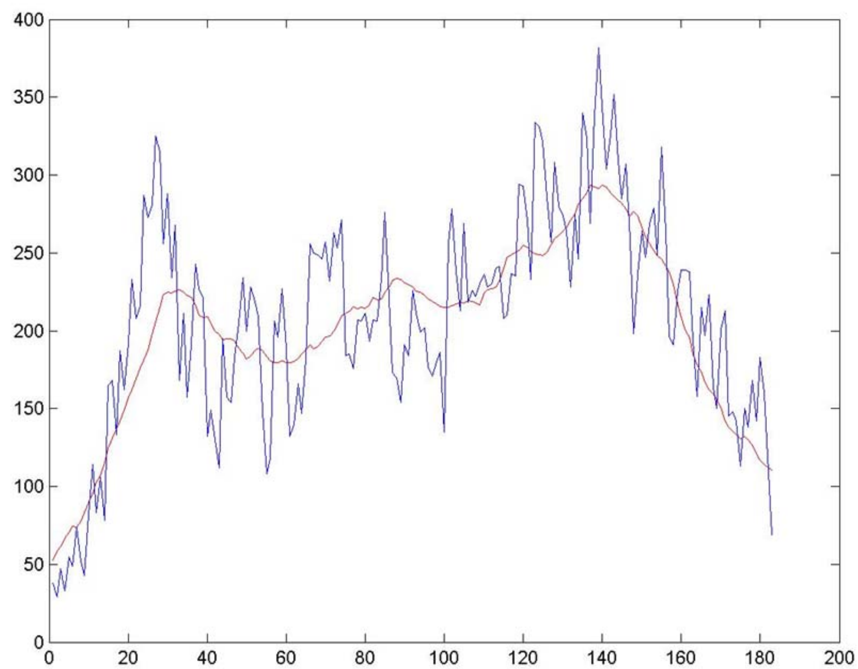
TABLE 1 Prediction Results

Prediction Horizon		Mall Area							Non-Mall Area						
		Eastbound				Westbound			Eastbound			Westbound			
	Link NO	3035	45	75	3225	76	46	36	165	215	275	266	276	216	166
1-step	RMSE	88.665	154.336	92.866	83.724	76.248	82.918	39.108	91.791	132.965	105.946	59.202	85.76	98.241	87.386
	MAPE	0.159	0.283	0.202	0.143	0.128	0.121	0.147	0.115	0.19	0.129	0.08	0.11	0.098	0.093
	Avg. MAPE (1)	0.197				0.132			0.145			0.095			
	Avg. MAPE (2)	0.169							0.116						
5-step	RMSE	89.219	158.438	96.341	85.125	78.39	83.957	39.261	95.13	138.739	109.201	61.138	90.171	103.202	92.16
	MAPE	0.16	0.292	0.21	0.146	0.132	0.123	0.148	0.119	0.199	0.134	0.083	0.117	0.103	0.099
	Avg. MAPE (1)	0.202				0.134			0.151			0.101			
	Avg. MAPE (2)	0.173							0.122						
15-step	RMSE	89.435	160.29	98.039	85.569	78.96	84.382	39.267	96.197	141.366	110.85	61.75	92.229	104.7	93.784
	MAPE	0.161	0.296	0.214	0.146	0.132	0.123	0.148	0.12	0.204	0.136	0.085	0.12	0.104	0.101
	Avg. MAPE (1)	0.204				0.134			0.153			0.103			
	Avg. MAPE (2)	0.174							0.124						
60-step	RMSE	89.448	160.515	98.243	85.595	78.984	84.406	39.268	96.277	141.583	110.985	61.776	92.324	104.759	93.843
	MAPE	0.161	0.296	0.215	0.147	0.132	0.123	0.148	0.121	0.205	0.136	0.085	0.12	0.105	0.101
	Avg. MAPE (1)	0.205				0.134			0.154			0.103			
	Avg. MAPE (2)	0.175							0.125						

Avg. MAPE : Average MAPE

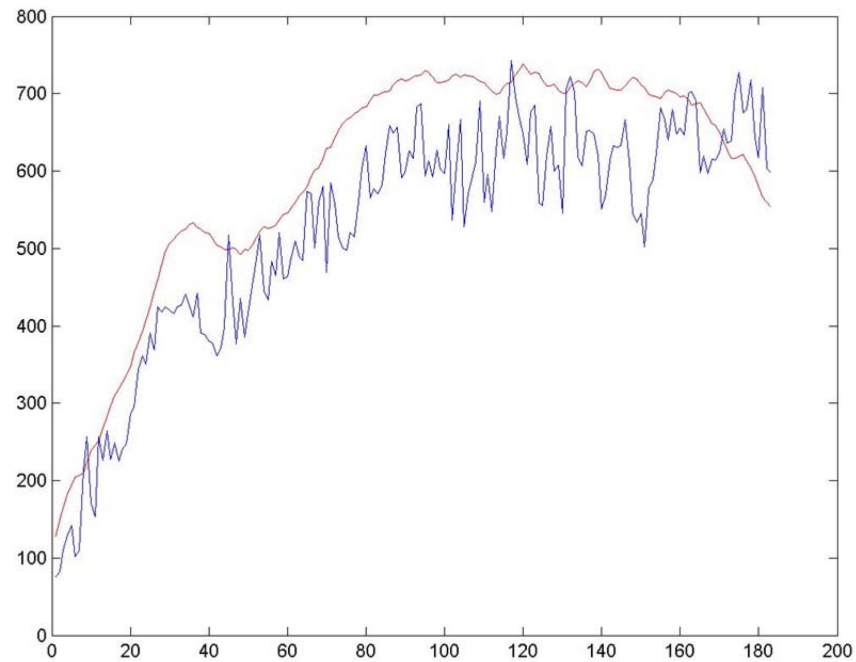


(a)

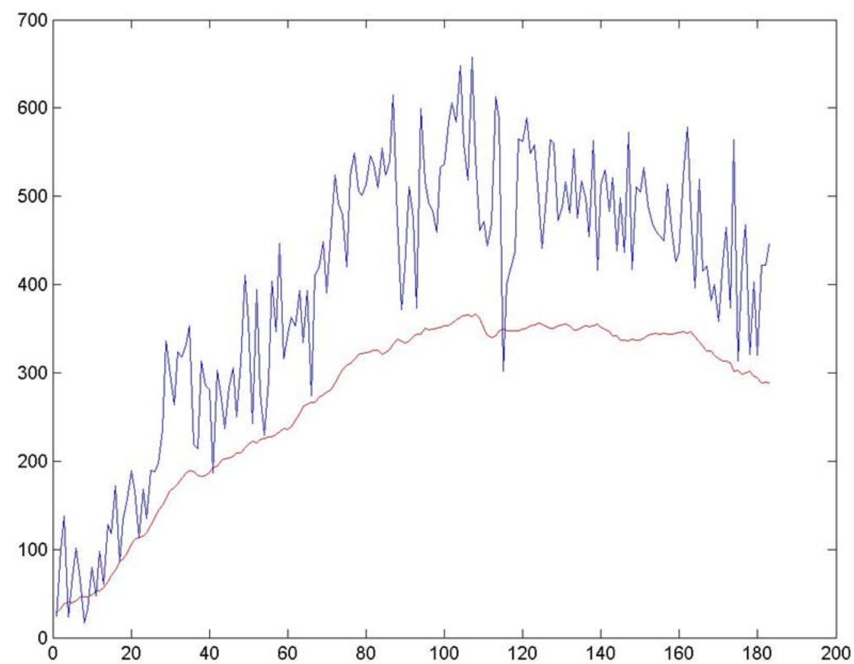


(b)

FIGURE 3 Examples of Prediction Results (a) July, 6, 2008 (Wed) (Link 165), and (b) July, 5, 2008 (Tue) (Link 36)



(a)



(b)

FIGURE 4 Links with low prediction accuracy , (a) July, 5th, 2008 (Tuesday) (Link 215) (b) July, 5th, 2008 (Tuesday) (link 45)

7. CONCLUSIONS AND RECOMMENDATIONS

Predicting traffic on an urban traffic network using spatio-temporal models has become a popular research area. The paper proposes a STRE model that can predict traffic volume by considering many detectors simultaneously. The City of Bellevue, Washington is selected as the test site because the City has more than 700 detectors covering the entire city. 105 detectors are included in the modeling process and the detectors on 8th Ave between a large shopping mall and freeway are used to demonstrate the prediction capability of the STRE model. This is because 8th Ave is considered one of the busiest streets in the City. The experiments show the STRE model can effectively predict traffic volume. Without further tune-up, all the experimental links have MAPEs between 8% and 15% except three special locations, Link 45 (Overall MAPE ≈ 29%), Link 75 (Overall MAPE ≈ 21%) and Link 215 (Overall MAPE ≈ 20%). As discussed, 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 *(9),* most other algorithms result in MAPEs ranging from 6% to 20%. Considering the high volatility of our test network and active interaction between each block, the STRE model is encouraging.

Even though the STRE model provides encouraging prediction results, many challenges still exist. Importantly, many parameters need to be adjusted during the calibration process. In the meantime, pre-knowledge of traffic patterns would facilitate the model-tuning process. For future model improvement, one can follow many potential directions. First, investigating how to decide the number of reference points and locations is an issue worth being addressed in the future. Second, the selection of the basis function is critical. Once the basis function is determined, the tune-up process is also challenging. For example, the proposed penalty function in the basis function might need to change. A case-by-case basis might tremendously improve the results, especially for Links 215 and 45 that underperform in the study. Only through-movement detectors are used in this study. If the turning-movement counts are available, the penalty value can be more precisely determined to increase prediction accuracy.

ACKNOWLEGMENTS

The authors would like to thank Mark Poch, Fred Liang, Mike Whiteaker and Brook Durant in the City of Bellevue for providing arterial traffic data. The authors are also grateful to the STAR Lab for maintaining the arterial database and providing the online portal to access the data. Special thanks to Ari Daniels for English consultation and editing.

REFERENCES

1. Min, W., and L. Wynter. Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies*, Vol. 19, No. 4, 2011, pp. 606-616.
2. Cheng, T., J. Haworth, and J. Wang. Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems*, 2011, pp. 1-25.
3. Wu, Y.-J., X. Ma, and Y. Wang. Development of a Web-based Arterial Network Analysis System for Real-time Decision Making In Transportation Research Board 90th Annual Meeting. CD-ROM. Transportation Research Board, 2011.
4. Smith, B.L., B.M. Williams, and R. Keith Oswald. Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, Vol. 10, No. 4, 2002, pp. 303-321
5. Smith, B.L., and M.J. Demetsky. Traffic Flow Forecasting: Comparison of Modeling Approaches. *Journal of Transportation Engineering*, Vol. 123, No. 4, 1997, pp. 261-266.
6. Williams, B.M., and L.A. Hoel. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results. *Journal of Transportation Engineering*, Vol. 129, No. 6, 2003, pp. 664-672.
7. Hamed, M.M., H.R. Al-Masaeid, and Z.M.B. Said. Short-Term Prediction of Traffic Volume in Urban Arterials. *Journal of Transportation Engineering*, Vol. 121, No. 3, 1995, pp. 249-254.
8. Smith, B.L., and R.K. Oswald. Meeting Real–Time Traffic Flow Forecasting Requirements with Imprecise Computations. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 18, No. 3, 2003, pp. 201-213.
9. Tan, M.-c., S.C. Wong, J.-M. Xu, Z.-R. Guan, and P. Zhang. An Aggregation Approach to Short-Term Traffic Flow Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, No. 1, 2009, pp. pp 60-69.
10. Thomas, T., W. Weijermars, and E. van Berkum. Predictions of Urban Volumes in Single Time Series. *Intelligent Transportation Systems, IEEE Transactions on*, Vol. 11, No. 1, 2010, pp. 71-80.
11. Kamarianakis, Y., and P. Prastacos. Forecasting Traffic Flow Conditions in an Urban Network: Comparison of Multivariate and Univariate Approaches. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1857, 2003, pp. 74-84.
12. Kamarianakis, Y., A. Kanas, and P. Prastacos. Modeling Traffic Volatility Dynamics in an Urban Network. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1923, 2005, pp. 18-27.
13. Stathopoulos, A., and M.G. Karlaftis. A multivariate state space approach for urban traffic flow modeling and prediction. *Transportation Research Part C: Emerging Technologies*, Vol. 11, No. 2, 2003, pp. 121-135.
14. Ghosh, B., B. Basu, and M. O'Mahony. Multivariate Short-Term Traffic Flow Forecasting Using Time-Series Analysis. *Intelligent Transportation Systems, IEEE Transactions on*, Vol. 10, No. 2, 2009, pp. 246-254.
15. Min, X., J. Hu, Q. Chen, T. Zhang, and Y. Zhang, Short-term traffic flow forecasting of urban network based on dynamic STARIMA model, *Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on*, 2009, pp. 1-6.

16. Min, X., J. Hu, and Z. Zhang, Urban traffic network modeling and short-term traffic flow forecasting based on GSTARIMA model, *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, 2010, pp. 1535-1540.
17. Vlahogianni, E.I., M.G. Karlaftis, and J.C. Golias. Spatio-Temporal Short-Term Urban Traffic Volume Forecasting Using Genetically Optimized Modular Networks. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 22, No. 5, 2007, pp. 317-325.
18. Vlahogianni, E.I. Enhancing Predictions in Signalized Arterials with Information on Short-Term Traffic Flow Dynamics. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, Vol. 13, No. 2, 2009, pp. 73 - 84.
19. Zou, H., Y. Yue, Q. Li, and Y. Shi, A spatial analysis approach for describing spatial pattern of urban traffic state, *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, 2010, pp. 557-562.
20. Cressie, N., T. Shi, and E.L. Kang. Fixed Rank Filtering for Spatio-Temporal Data. *Journal of Computational and Graphical Statistics*, Vol. 19, No. 3, 2010, pp. 724-745.
21. Katzfuss, M., and N. Cressie. Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, Vol. 32, No. 4, 2011, pp. 430–446.
22. Cressie, N., and C.K. Wikle. Space–time Kalman filter. *Encyclopedia of Environmetrics*, Vol. 4, 2002, pp. 2045–2049.
23. Ma, X., Y.-J. Wu, and Y. Wang. DRIVE Net: An E-Science of Transportation Platform for Data Sharing, Visualization, Modeling, and Analysis In Transportation Research Board 90th Annual Meeting. CD-ROM. Transportation Research Board, 2011.
24. Xie, Y., Y. Zhang, and Z. Ye. Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 22, No. 5, 2007, pp. 326-334.
25. Washington, S., M.G. Karlaftis, and F.L. Mannering, Statistical and Econometric Methods for Transportation Data Analysis, Chapman & Hall/CRC, Boca Raton, 2003.