KSCE Journal of Civil Engineering (2023) 27(7):3074-3084
DOI 10.1007/s12205-023-1339-7

pISSN 1226-7988, eISSN 1976-3808
www.springer.com/12205

Transportation Engineering

Short-Term Traffic Speed Prediction for Multiple Road Segments

Bumjoon Bae[a] and **Lee D. Han**[b]

[a]Member, Center for Privately-Financed Highway Studies, Korea Transport Institute, Sejong 30147, Korea
[b]Dept. of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996, USA

ARTICLE HISTORY

Received 22 August 2022
Revised 9 March 2023
Accepted 26 April 2023
Published Online 29 May 2023

KEYWORDS

Traffic speed prediction
Principal component analysis
Singular spectrum analysis
Vector autoregressive model
Spatiotemporal dependency

ABSTRACT

Short-term traffic prediction has been an essential part of real-time applications in modern transportation systems for the last few decades. Despite the recent progress in the voluminous models and data sources, many existing studies have focused on prediction for either a single or a few locations. In addition, the spatiotemporal dependency in the traffic data was narrowly accounted for. Therefore, this paper finds a new short-term traffic speed prediction algorithm that can efficiently cope with the complexity and immensity of the prediction process derived from the network size and amount of data in order to provide accurate predictions in real time. This algorithm consists of two modules: (a) principal component analysis (PCA) for data dimensionality reduction and feature selection, and (b) multichannel singular spectral analysis (MSSA) for multivariate time-series data prediction. A large amount of traffic data is efficiently compressed by PCA with high accuracy, then used as an input in the multivariate time-series analysis. The algorithm was compared with a vector autoregressive (VAR) model to predict traffic speeds five minutes ahead for a 21.3-mile-long highway segment, using the traffic detector data, and for 451-mile-long segment, using probe-based speed data in Tennessee. The tested algorithm is found to provide accurate predictions with a computation time of less than one second without training. Furthermore, the algorithm shows a better prediction performance under congested flow conditions, compared to VAR. This indicates that the tested algorithm is suitable for real-time prediction and scalable for a large network analysis.

1. Introduction

Traffic speed is one of the fundamental variables that characterize traffic flow. It is not only a traffic performance measurement of roadway systems, but also an input for estimating other measurements such as travel time, vehicle emission, traffic noise, and so on (May, 1990). Hence, traffic speed prediction is a core function required in modern traffic management and operation systems. In the last few decades, various short-term traffic speed prediction models and algorithms have been developed for real-time intelligent transportation systems (ITS) applications.

Although there is no absolute definition of how long the ‘short-term’ is, the prediction time step varies from 30 seconds to 5 minutes in the literature (Alecsandru and Ishak, 2004; Ishak and Alecsandru, 2004; Vanajakshi and Rilett, 2004; Yang et al., 2004; Chandra and Al-Deek, 2008; Chandra and Al-Deek, 2009; Guo and Williams, 2010; Min and Wynter, 2011;Ye et al., 2012). And the prediction horizon has been set as the range from one minute to two hours in advance through multi-step runs (Vlahogianni et al., 2014). According to a recent comprehensive review on short-term traffic forecasting by Vlahogianni et al. (2014), the majority of the previous studies used univariate models with traffic detector data at a single location on a highway. Although a few studies utilize a multi-source traffic speed data, they do not focus on forecasting but on prediction for unobserved links or locations (Bae et al., 2018; Lin et al., 2018). Statistical time-series models and neural network (NN) type models present a noticeable frequency of use. The time-series models include vector autoregressive (VAR) models for multivariate prediction (Chandra and Al-Deek, 2008; Chandra and Al-Deek, 2009), spatial temporal autoregressive moving average (STARMA) for considering spatiotemporal correlation (Min and Wynter, 2011), generalized autoregressive conditional heteroscedasticity (GARCH) for capturing unexpected speed dynamic shifts (Guo and Williams, 2010), and adaptive Lasso regression for improving prediction performance by minimizing error variance (Kamarianakis et al.,

CORRESPONDENCE Bumjoon Bae bbae2016@koti.re.kr Center for Privately-Financed Highway Studies, Korea Transport Institute, Sejong 30147, Korea

2012), and so on. On the other hand, a variety of NN based models has also been proposed for speed prediction. These models are known to provide a more accurate prediction for nonlinear traffic flow compared to the classical statistics models (Amin et al., 1998; Dia, 2001; Quek et al., 2006; Chan et al., 2013). These models have also been tested with Kalman filters or wavelet transformation technique primarily for denoising traffic data (Yang et al., 2004; Heilmann et al., 2011; Wang and Shi, 2013). Recent studies test deep learning models for traffic flow prediction and show that such models provide satisfactory results for non-recurring congestion with certain data input conditions (Chan et al., 2013; Lv et al., 2015; Polson and Sokolov, 2017; Liu et al., 2020). Another type of speed prediction is using macroscopic traffic flow models, such as FREFLOW (Payne, 1971), KRONOS (Michalopoulos et al., 1993), and METANET (Papageorgiou et al., 1990), focusing on the aggregated traffic flow characteristics represented by speed, flow and density (Fang and Jin, 2014). Although their underlying model captures spatiotemporal information, it utilizes traffic flow characteristics of a few adjacent road links and time interval. A recent comparable study, Fang and Jin (2014) shows that the prediction error levels of METANET-based models were relatively higher than those of statistical and NN models.

In the literature, the mean absolute prediction error (MAPE) of existing studies ranges from 2.5% to 15.0% for five-minute predictions (Alecsandru and Ishak, 2004; Ishak and Alecsandru, 2004; Yang et al., 2004; Chandra and Al-Deek, 2008; Min and Wynter, 2011; Kamarianakis et al., 2012; Chan et al., 2013; Wang and Shi, 2013). Although the effects of variability in the time step on prediction performance has not been addressed sufficiently, the prediction error shows generally a linear association with the length of a prediction time step or the number of time steps increase (Alecsandru and Ishak, 2004; Ishak and Alecsandru, 2004; Djuric, et al., 2011; Min and Wynter, 2011; Kamarianakis et al., 2012). A few studies compared the prediction performances of congested and non-congested traffic flow conditions. They showed that the prediction errors of congested conditions are approximately three times higher than those of non-congested conditions (Chandra and Al-Deek, 2008; Guo and Williams, 2010). The speed threshold to define congestion varies over the studies, ranging from 30 miles per hour (mph) to 40 mph.

Despite the extensive studies on short-term traffic speed prediction, few have attempted to address the following limitations. The existing studies applied five minutes as a prediction time step without considering its effects. This is mainly because five minutes had been used most frequently in literature and the available data resolution was five minutes. Furthermore, there was insufficient information in the literature on computation time evaluation as real-time applications, which is helpful for other researchers and practitioners. In addition, many of the previous studies have been done on the short-term prediction for a single or several locations, in which a spatiotemporal dependency of traffic data was not sufficiently considered.

This paper tests a new short-term traffic speed prediction algorithm for multiple road segments. To support real-time and proactive traffic operations, the tested algorithm aims to predict the future traffic flow conditions accurately and quickly without training a model. It is a data-adaptive algorithm that can handle a large-scale spatiotemporal speed data within a short amount of time.

The remainder of this chapter is in this manner. The next section details the methodologies used in the tested algorithm. Then, the data sources and different aspects of testing performance are described. Next, the prediction performance of the algorithm is compared with that of vector autoregressive (VAR) model that has been successful in the past 10 years (Chandra and Al-Deek, 2008; Chandra and Al-Deek, 2009; Vlahogianni et al., 2014). Finally, a discussion on the results and conclusion are drawn.

2. Methodology

The proposed algorithm consists of principal component analysis (PCA) and multichannel singular spectrum analysis (MSSA) (see Fig. 1). First, PCA is used to extract features and reduce dimensions of the data. Then, MSSA is used for multivariate time-series prediction using the principal components from PCA. This approach has achieved satisfactory performance in medical image processing studies (Mizuguchi et al., 2010; Chhatkuli et al., 2015).

Unlike the statistical prediction models such as autoregressive integrated moving average (ARIMA), MSSA, a multivariate extension of singular spectrum analysis (SSA) is a data-adaptive time-series analysis method that does not require any assumptions, such as stationarity of the data, linearity of the model, or normality of the residuals (Hassani et al., 2013; Hassani et al., 2015). These features make MSSA useful (Elsner and Tsonis, 1996; Patterson et al., 2011; Hassani et al., 2013; Alessio, 2016). Hence, SSA and MSSA have been widely applied recently in many disciplines such as economics, medical image processing, climatology research, etc. (Vitanov et al., 2008; Mizuguchi et al., 2010; Cressie and Wikle, 2011). More theoretical and mathematical details of SSA can be found in (Elsner and Tsonis, 1996) and (Hassani and Thomakos, 2010). Furthermore, using the principal components (PC) as an input of MSSA allows the prediction to be made based on spatiotemporal dependencies in the data. According to

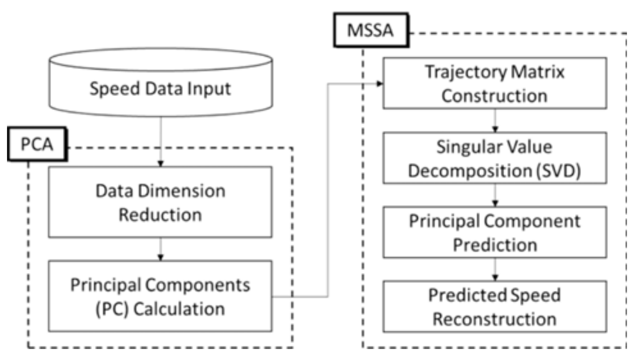


Fig. 1. Proposed Speed Prediction Algorithm

Asif et al. (2013), PCA consistently provides high reconstruction accuracy over different compression rates for spatiotemporal traffic data.

To simultaneously predict the traffic speeds of multiple road segments, the proposed algorithm requires multivariate time-series traffic speed data as an input. Through the first part of the algorithm, the PCA module, the input data matrix is orthogonally transformed to reduce its size, keeping the essential features. Then, the transformed matrix is used as an intermediate input for the second part, MSSA. During the MSSA module process, the multivariate feature data matrix is first decomposed to extract important features except for noise or outliers. Then, the predicted data for the future time interval is derived through the optimization problem. Finally, the predicted data is reconstructed to the original multivariate time-series format. More details of the algorithm are explained as follows.

2.1 Principal Component Analysis

Principal component analysis (PCA) is a widely used multivariate statistical procedure used for data dimension reduction and feature extraction (Chen et al., 2009). It is an orthogonal transformation method that projects the original data onto the spaces of linearly uncorrelated variables where the variance is maximized based on eigenvalues and eigenvectors. Therefore, the principal components (PC), the transformed data can be used as an input for a variety of post analyses.

The speed observation x_{it}, $1 \le i \le n$, $1 \le t \le p$, with i representing location and t representing time, gives the multivariate time-series data as

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}, \tag{1}$$

The covariance matrix is calculated as

$$C = \frac{1}{p}\sum_{t=1}^{p} \Psi_t \Psi^T = \Phi\Phi^T , \tag{2}$$

where, $\Psi_t = X_t - \mu$, which is the vector difference between the observations at time t and the mean of X, μ. Since $\Phi = \frac{1}{\sqrt{p}}[\Psi_1, \ldots, \Psi_p]$, the dimension of the covariance matrix C is $(n \times n)$.

As the road network size to be analyzed is increased, especially when $n >> p$, calculating $\Phi\Phi^T$ and its eigenvectors becomes more intractable. In order to near-real-time analysis, Turk and Pentland (1991) proposed to use $\Phi^T\Phi$ instead of $\Phi\Phi^T$, which reduces the dimension from $(n \times n)$ to $(p \times p)$. This approach is very common in image processing analysis where the input data at each time step is usually a 2-dimentional image. For example, if the input data size is $(n \times n)$, the size of time-series data, X is $(n^2 \times p)$, so $\Phi\Phi^T$ gives $(n^2 \times n^2)$ covariance matrix. More details about the relationship of $\Phi^T\Phi$ with $\Phi\Phi^T$ is provided in (3) through (6).

The eigenvector v_i is defined as

$$\Phi^T\Phi v_i = \lambda_i v_i , \tag{3}$$

where, λ_i is the eigenvalue of $\Phi^T\Phi$ denoted by $\lambda_1 \ge \cdots \ge \lambda_p$. If Φ is multiplied in both sides of (3),

$$\Phi\Phi^T\Phi v_i = \lambda_i \Phi v_i , \tag{4}$$

and using (2) and (4),

$$C\Phi v_i = \lambda_i \Phi v_i . \tag{5}$$

Then, (5) can be expressed as

$$C u_i = \lambda_i u_i . \tag{6}$$

Therefore, $\Phi\Phi^T$ and $\Phi^T\Phi$ have the same eigenvalues and their eigenvectors have the relationship as $u_i = \Phi v_i$.

Finally, the orthogonally transformed data, Y is computed by using the $(p \times p)$ eigenvectors, u as follows.

$$Y = X^T u \tag{7}$$

The resultant $(p \times p)$ matrix, Y from (7) is used as an input data for the following MSSA procedure.

2.2 Multichannel Singular Spectrum Analysis

The first step of MSSA is called embedding, which means mapping each univariate time series into multivariate series using subsets of the univariate time series. This procedure is similar to a time series analysis based on moving average calculation (Patterson et al., 2011). For example, using the kth column of Y, $[y_1^{(k)}, y_2^{(k)}, \ldots, y_p^{(k)}]$, the resultant matrix of embedding, called trajectory matrix, is defined as

$$y^{(k)} = \begin{bmatrix} y_M^{(k)} & y_{M+1}^{(k)} & \cdots & y_p^{(k)} \\ y_{M-1}^{(k)} & y_M^{(k)} & \cdots & y_{p-1}^{(k)} \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(k)} & y_2^{(k)} & \cdots & y_{p-M+1}^{(k)} \end{bmatrix}, \tag{8}$$

where, M is the embedding dimension (also called window length) which is an arbitrary integer that $2 \le M \le p$. Alessio (2016) provides a "reasonable" range of M that is greater than the number of data points in which one oscillatory pattern to be detected and less than $p/5$. However, it is better to choose the value of M based on the comparison of the results from different values of M. Therefore, a sensitivity analysis was conducted in the case study to investigate the effects of choosing the values of p and M in the next chapter.

$y_{cr}^{(k)}$ is the centered matrix of $y^{(k)}$ based on each row mean, calculated as.

$$y_{cr}^{(k)} = y_{cr}^{(k)} - \frac{1}{p-M+1}\begin{bmatrix} y_M^{(k)} + y_{M+1}^{(k)} + \cdots + y_p^{(k)} \\ y_{M-1}^{(k)} + y_M^{(k)} + \cdots + y_{p-1}^{(k)} \\ \vdots \\ y_1^{(k)} + y_2^{(k)} + \cdots + y_{p-M+1}^{(k)} \end{bmatrix}. \tag{9}$$

Then, the trajectory matrix of MSSA is made as

$$Y_{cr} = \begin{bmatrix} y_{cr}^{(1)} \\ y_{cr}^{(2)} \\ \vdots \\ y_{cr}^{(K)} \end{bmatrix}, \tag{10}$$

where K is the number of selected PCs corresponding to the Kth largest eigenvalues in (6) $(1 \le k \le K)$. Y_{cr} is a $(KM \times p')$ matrix and $p' = p - M + 1$. What to be estimated is the next column (i.e., $p - M + 2$ th column) of Y_{cr}. This is defined as

$$Z = [y_{cr,p+1}^{(1)}, y_{cr,p}^{(1)}, \ldots, y_{cr,p-M+2}^{(1)}, \ldots, y_{cr,p+1}^{(K)}, y_{cr,p}^{(K)}, \ldots, y_{cr,p-M+2}^{(K)}]\,. \tag{11}$$

In this study the number of PCs from (7) is selected to make losing information of the original data, X by less than 1%. Resultantly, the selected PCs explain over 99.7% of the total variance of the data in the case study. The percentage can be calculated as the proportion of the sum of the K largest eigenvalues over the sum of all eigenvalues of X. In MSSA, the row length of matrix Z gets longer as the road network size increases, compare to SSA. Then, the dimension becomes much larger after being squared in the following step. Fig. 2 shows the percentage of data dimension reduction by using PCA for MSSA. Compare to the case of using MSSA without PCA, for example, the data dimension in MSSA is reduced by approximately 90% by PCA if the original data dimension of $(n \times p)$ is (300×100). A different number of PCs can be selected by employing information criteria, such as AIC, ICOMP, etc.

The next step of MSSA is a singular value decomposition (SVD) of the squared trajectory matrix, $C_Y = Y_{cr}Y_{cr}^T$. The elements of the lagged-covariance matrix C_Y reflect the linear correlation between the all pair of patterns in the embedding window. Thus, the recurring patterns in the time series result in a relatively high covariance in C_Y (Elsner and Tsonis, 1996). Through SVD, C_Y is decomposed into orthogonal eigenvectors as follows.

$$C_Y = E\Lambda E^T, \tag{12}$$

where, E is the eigenvectors of C_Y which are the singular vectors

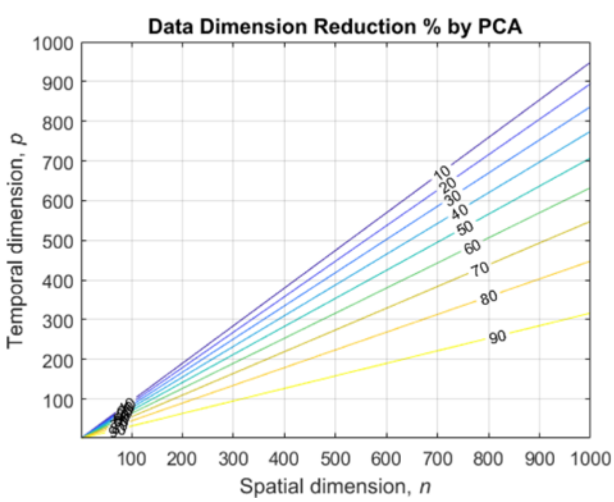


Fig. 2. Data Dimension Reduction Rate of MSSA by Using PCA

of Y_{cr}, and Λ is a diagonal matrix that consists of ordered values, equal or greater than zero, whose square roots are the singular values of Y_{cr}. Then, the L largest eigenvalues from Λ and corresponding eigenvectors from E are selected for prediction as (13). In this study $L = p$ is applied which is large enough to contain the most significant eigenvectors. Through this step, the recurring patterns in the time series can be separate and the noise in the data can be removed (Patterson et al., 2011).

$$W = [E^{(1)}, E^{(2)}, \ldots, E^{(L)}]\,. \tag{13}$$

Using the selected $(KM \times L)$ eigenvector matrix W, the estimation of Z is given as the least-squares problem as follows (Hassani and Thomakos, 2010; Mizuguchi et al., 2010; Chhatkuli et al., 2015).

$$\text{minimize } (Z - WW^TZ)^2 \tag{14}$$

This implies that the evolution of the next vector in the trajectory matrix follows the same law of the other adjacent vectors (Loskutov et al., 2001).

Then, Z can be decomposed as,

$$Z = RP + Q \tag{15}$$

where $P = [y_{cr,p+1}^{(1)}, y_{cr,p+1}^{(2)}, \ldots, y_{cr,p+1}^{(K)}]^T$. The $(KM \times K)$ and $(KM \times 1)$ restriction matrices, R and Q are defined as follows.

$$R = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \vdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \vdots & 1 & \cdots & \vdots \\ \vdots & 0 & \cdots & \vdots \\ \vdots & \vdots & \cdots & 1 \\ \vdots & \vdots & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix} \tag{16}$$

$$Q = [0, y_{cr,p}^{(1)}, \ldots, y_{cr,p-M+2}^{(1)}, \ldots, 0, y_{cr,p}^{(K)}, \ldots, y_{cr,p-M+2}^{(K)}]\,,$$

By decomposing (14) with (15), the future component of the time series data can be obtained as (17) (Mizuguchi et al., 2010; Chhatkuli et al., 2015).

$$P = (I - R^TWW^TR)^{-1}R^TWW^TQ\,, \tag{17}$$

where, I is a $(K \times K)$ identity matrix.

Finally, the predicted speed is calculated by re-centering the values of P and multiplying them with the eigenvectors from (6). The final step is necessary because the predicted values of P are orthogonally transformed and centered during prior steps.

3. Case Study

3.1 Data Description

The proposed prediction algorithm was applied to speed data for Interstate 40 (I-40) in Tennessee from two data sources: (a) traffic detector data, named Remote Traffic Microwave Sensors (RTMS), which is collected every 30 seconds from over 1,000 traffic detector stations on interstate highways in Tennessee, and (b) probe-based link speed data, named National Performance Management Research Data Set (NPMRDS). For RTMS, the detector stations are located only in major urban areas of the state. Therefore, 41 stations in the 21.3 mile-long westbound I-40 segment were selected, which is a major corridor in Knoxville, Tennessee. The stations are on average 0.5 miles from each other. Traffic speeds for the intermediate locations in 0.1-mile increments between two consecutive stations were interpolated using the adaptive smoothing method (Treiber et al., 2011) in order to augment the spatial resolution of the data by 213.

The speed data from September 23 and September 30 in 2016, both of which were Fridays, were collected from the detectors and averaged in five minutes, i.e., the data dimension is (213 × 288) for each day. Both days were selected based on the fact that there was no incident in the first day whereas there was a severe incident on the second day. The incident was verified by the traffic incident data log from the local transportation management center (TMC). Since prediction of unexpected events, such as crashes, adverse weather conditions, etc., in the spatiotemporal domain is highly intractable, it is worth testing how quickly the speed prediction algorithm can adapt or how sensitive it is to sudden changes in traffic conditions.

In order to evaluate the proposed algorithm performance for a longer road segment, i.e., larger data dimension, the NPMRDS data were used. For NPMRDS, the spatial coverage is the entire interstate highway systems in the state. In this study, the five-minute average speeds of NPMRDS for the 298 road links of a 451-mile-long I-40 westbound segment on February 3rd, 2017 were collected. Please note that five minutes are the highest resolution for the available NPMRDS dataset, i.e., the data dimension is (298 × 288). Fig. 3 shows examples of the data visualizations.

3.2 Performance Measures

To evaluate the prediction performance of the proposed algorithm, two error measures were used, which are the mean absolute error (MAE) and mean absolute percentage error (MAPE). They are defined as follows.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|x_i - \hat{x}_i\right|, \tag{17}$$

$$PE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|x_i - \hat{x}_i\right|}{x_i} \times 100, \tag{18}$$

where, x_i is the observed traffic speed and $\hat{x}_i$ it the predicted traffic speed.

3.3 Data Resolution Selection

To choose an optimal prediction interval is an important issue which depends on the type of ITS applications, algorithms and data sources (Vlahogianni et al., 2014). In order to investigate the effect of the data resolution on the short-term traffic speed prediction for the proposed algorithm, as a similar approach in Guo et al. (2007), a sensitivity analysis framework was applied. The need for a sensitivity analysis is mainly due to the nonparametric characteristic of PCA-MSSA, i.e., it does not allow to test statistical significance of parameter estimates. Four datasets of the 24-hour traffic speeds from RTMS were generated by different aggregation levels: 0.5-, 1-, 2.5-, and 5-minute and used in a preliminary analysis. To make predictions for the target time in the future, the iterative predictions are made, i.e., the predicted values are added to the initial data for the next prediction. For selecting the temporal aggregation scale, the target time to predict was set as 5 minutes, observed most frequently in the literature. Guo et al. (2007) also recommended 5 minutes as the shortest interval, although the investigations and findings were for an online prediction algorithm based on a SARIMA + GARCH structure. Table 1 shows the average prediction performance for 5-minute prediction. Each prediction was made using the past thirty data points. To predict the next five-minute traffic speed, for example, the prediction process is implemented ten times iteratively using the 30-second dataset. As the number of prediction steps increases, the prediction error increases. This is because the error in the current prediction is transferred to the next prediction step. Therefore, five minutes gave the lowest errors for the five-minute prediction. This was expected because the increase in the data resolution reduces the inherent noise in the data making the series more stable. And the result is consistent with the findings in the literature (Guo et al., 2007).

Table 1. Temporal Scale Effects on 5-minute Prediction Performance Using RTMS

MOEs	Data resolution (Number of prediction steps)			
	0.5 min (10)	1 min (5)	2.5 min (2)	5 min (1)
MAE (mph)	3.40	3.31	3.12	3.03
MAPE (%)	9.67	9.16	8.23	7.94

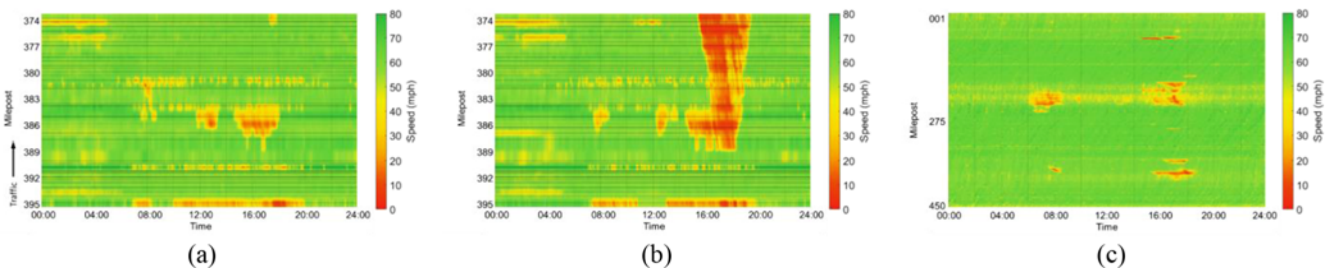


Fig. 3. Speed Data Visualization: (a) RTMS – September 23, 2016, (b) RTMS – September 30, 2016, (c) NPMRDS – February 3, 2017

The following analyses were made using the data aggregated in five minutes.

3.4 Input Data Dimension and Window Length Selection

The effects of choosing different data length p and window length M were investigated in a sensitivity analysis. Here the range of 0.5-6 hours for both p and M was considered using the 5-minute RTMS data of September 23 and September 30 in 2016 and NPMRDS data on February 3, 2017. In order to choose proper values of p and M, MAPE and computation time for one-step prediction were compared as shown in Fig. 4. Please note that the vertical axis of the figures of MAPE in Fig. 4 represents 1/MAPE for better recognition of the best result. Fig. 4(a) shows that there is a gradual increase in MAPE with increase of both of p and M in the range of 1-5.5 hours. The computation time of Fig. 4(a) also shows the same pattern of the MAPE figure; however, it increases much more rapidly as p and M get closer to six hours. Similar patterns were observed in Fig. 4(b). Based on these sensitivity results, $p = 18$ (1.5 hours) and $M = 12$ (1 hours) for RTMS – September 23, 2016, $p = 24$ (2 hours) and $M = 18$ (1.5 hours) for RTMS – September 30, 2016, and $p = 36$ (3 hours) and $M = 18$ (1.5 hours) for NPMRDS were applied.

3.5 Prediction Performance

3.5.1 Comparison with a Benchmark Model

As Tan et al. (2016) mentioned, it is difficult to comparing prediction performance of different models because their objectives, spatiotemporal scope, and input data conditions can differ each other. Since this paper targets at higher prediction accuracy and shorter process time for multiple roadway links, a VAR model was selected as a benchmark to evaluate the performance of the proposed algorithm.

The speed prediction results for the next five minutes were compared to those of VAR(k). In this study, the order of the model k was determined to be within the range of 1-8 (i.e., $k = 1, ..., 8$) based on the goodness of fit of the model using Akaike's information criterion (AIC) (Akaike, 1973). The 24-hour historical speed data were used for each prediction target time point to train the VAR(k) model. The RTMS dataset was used to make 288 predictions for September 23 and 30, 2016. In order to compare the computation time, both methods were implemented on the same platform with Intel® Core™ i7 processor (3.60 GHz) with 8GB memory.

In this study, restricted VAR models were used. Unrestricted VAR models using a full covariance matrix for parameter estimation is not suitable for real- time data analysis on a large-scale network for these reasons: First, the model estimation time is too long because a large number of parameters will be estimated. For example, an unrestricted VAR(1) model with n = 213 has 68,373 ($= n + nAR{\cdot}n^2 + n(n+1)/2$, where the number of the autoregressive matrix, $nAR = 1$) parameters to be estimated, whereas a restricted model has only 639 ($=3n$). Therefore,

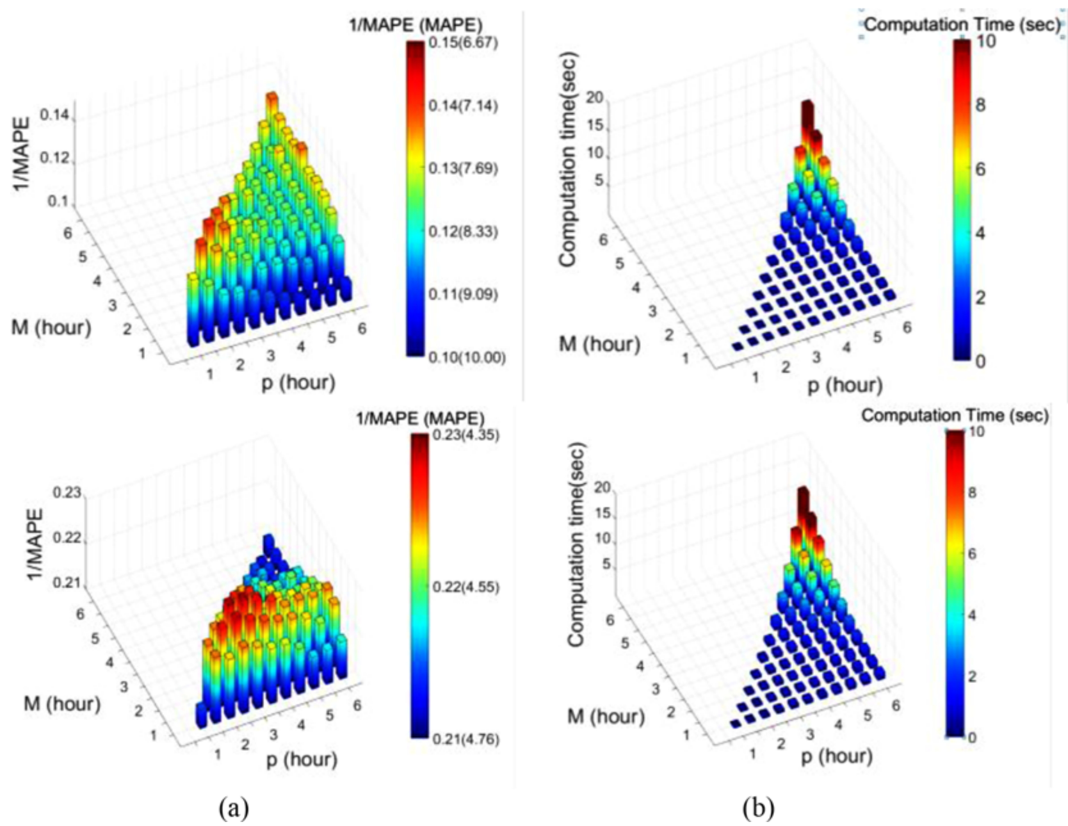


Fig. 4. MAPE (figures in the left column) and Computation Time (in the right column) for Different Temporal Dimension and Window Length: (a) RTMS – September 30, 2016; (b) NPMRDS – February 3, 2017

estimating an unrestricted VAR model takes too long when either the network n or the autoregressive lag k is large. Second, the residual process of the unrestricted model is likely to have a non-positive definite covariance matrix which makes parameter estimation impossible.

In order to investigate the effect of applying PCA in the proposed algorithm, MSSA without PCA, referred to hereafter as MSSA, was also tested. In addition, based on the fact that it is more likely to use a pre-trained model in practice, the VAR(k) model was separated into two types: (a) a model whose parameter estimates are updated for each prediction, denoted as On-VAR(k); and (b) a model whose parameter values are fixed once the model is trained priorly, denoted as Off-VAR(k). Please note that the model order k of On-VAR(k) is not updated for each prediction step; otherwise, training a model takes an excessive amount of time, making short-term prediction harder to achieve. Therefore, the same order k of Off-VAR(k) was applied to On-VAR(k). For the same reason, the On-VAR(k) model was trained using five-hour historical data for each prediction target time.

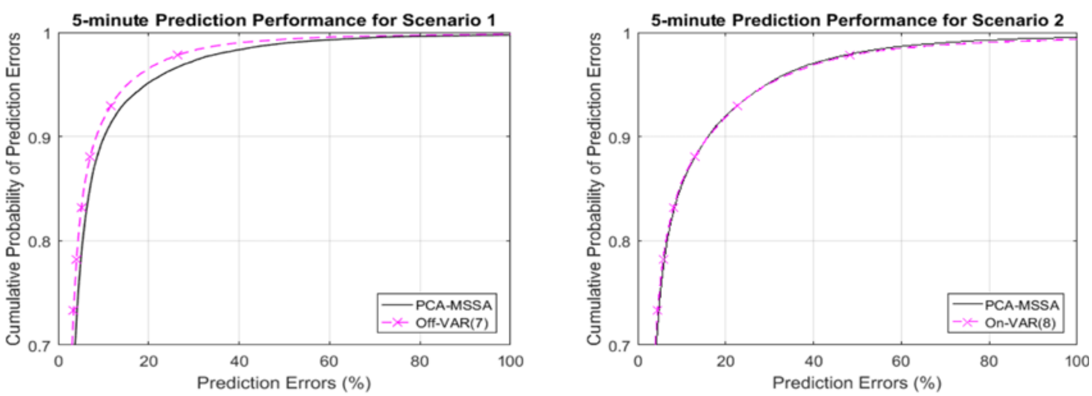
Table 2 summarizes the 5-minute prediction performances of these four methods. For Scenario 1 non-incident condition, Off-VAR(7) outperforms the others. In this scenario, traffic flow is very stable in terms of speed except for the congestion around milepost 386 during afternoon peak hours. For such cases, the speed data hold high stationarity and the VAR models fit the data well. The error level of PCA-MSSA is slightly higher than both VAR models and MSSA. In comparison with Off-VAR(7), as depicted in Fig. 5(a), the level of error of PCA-MSSA is slightly higher than that of Off-VAR(7) across the overall error range. This may result from the information loss of data in the dimension reduction procedure or the misspecified length of the input data and embedding window.

Despite the different prediction performance in Scenario 1, traffic prediction for a free-flow condition is not challenging. In other words, prediction of traffic conditions during the transitions to and from congested flow over time and space should be paid attention more. The traffic condition in Scenario 2 shows such instability in the speed data caused by a severe incident. As shown in Table 2, MSSA and PCA-MSSA outperform both VAR models. The MAPE of 6.40% from MSSA is slightly better than 6.56% from PCA-MSSA. Since the same dimension of input data was employed, it is probable that the different performance was caused by PCA. Contrary to the result in Scenario 1, On-VAR outperformed Off-VAR in Scenario 2 in terms of MAPE. On-VAR model predicts the congested flow better than Off-VAR by updating parameter estimates for each prediction. In order to evaluate the performance of PCA-MSSA for congested traffic flow, its prediction error range is compared with that of On-VAR in Fig. 5(b). Although the cumulative probability error curves of both methods are very similar, they intersect at around 25%. This indicates that the average error level of PCA-MSSA is relatively lower for low speed conditions, compared to On-VAR.

Figure 6 shows the predicted speed profiles of four methods at selected locations. The Fig. 6(a) location is in a weaving section where two major interstate highways are merged. Recurrent afternoon congestion was intensified due to an incident that

Table 2. Comparison of 5-minute Prediction Performance for RTMS

		PCA-MSSA	MSSA	On-VAR(k)	Off-VAR(k)
Scenario 1	Model	p=18, M=12	p=18, M=12	k=7	k=7
	MAE (mph)	2.31	2.26	2.19	1.96
	MAPE (%)	4.98	4.90	4.76	4.26
Scenario 2	Model	p=24, M=18	p=24, M=18	k=8	k=8
	MAE (mph)	2.46	2.39	2.52	2.41
	MAPE (%)	6.56	6.40	7.02	7.26
Average computation time (s)		0.05	6.78	114.20	0.22



(a) (b)

Fig. 5. Prediction Performance of PCA-MSSA and VAR: (a) Scenario 1, (b) Scenario 2

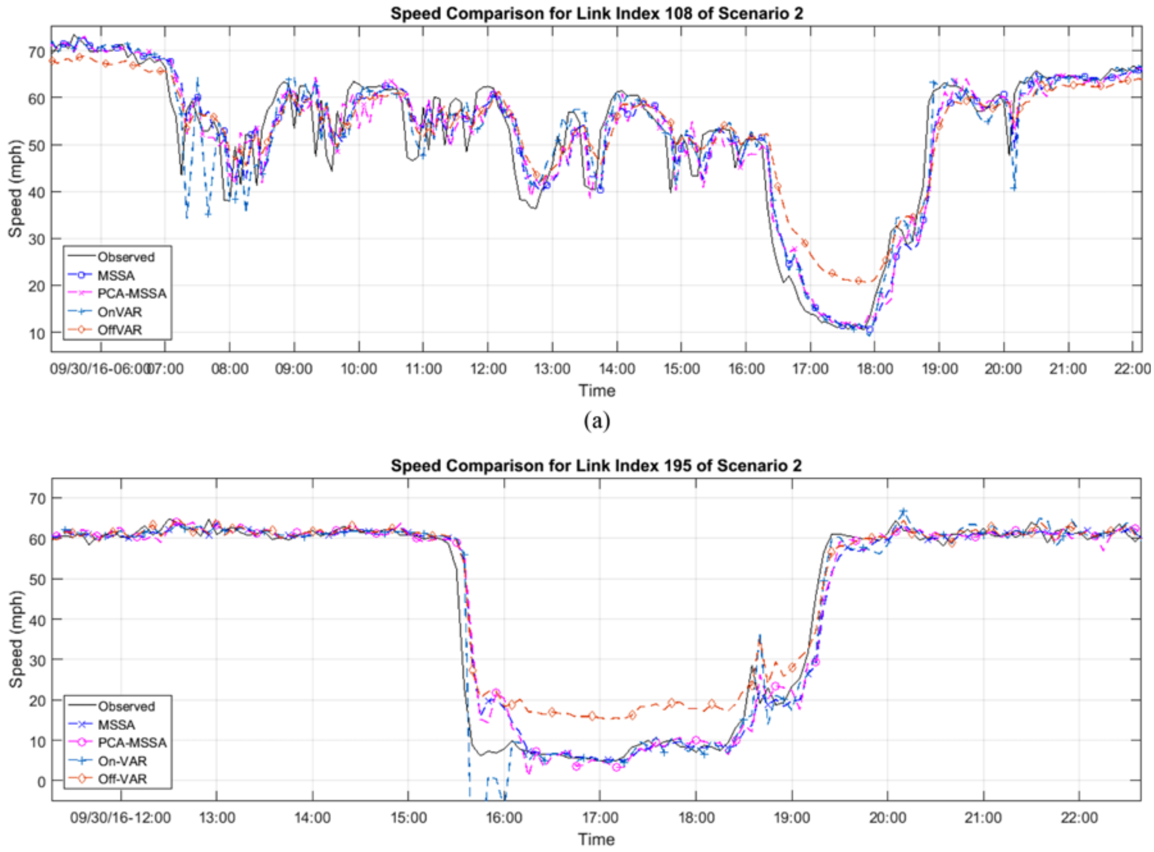


Fig. 6. Predicted Speed Profiles: (a) Location Index 108, (b) Location Index 195

occurred downstream around 3:00 to 4:00 PM. All the predicted profiles, except for Off-VAR, show similar patterns and follow the observed speed fluctuation. However, On-VAR tends to produce overfitted results when the traffic state changes from free-flow to congestion in the morning peak hours. The same pattern of On-VAR is also present in Fig. 6(b). Such overfitting patterns of VAR under a sudden change in traffic flow can also be observed in Polson and Sokolov (2017). The average performance measurement in space is shown in Fig. 7. Both PCA-MSSA and MSSA outperform the VAR models during the congested time period. With the emergence of congested traffic flow, all the error measures are increased. However, both MSSA algorithms quickly adapt to the changes of flow states so that their error measures are decreased. This is consistent with literature that showed SSA is suitable for time series with various structures, e.g., stationarity or non-stationarity, cyclical patterns or sharp edges, without losing information about important features of time-series data (Shang et al., 2016; Suksiri et al., 2016). As mentioned in Section 2, this is mainly because, unlike parametric models SSA does not require any assumptions of data but utilizes elementary signals separated from noise. Although the error of On-VAR also decreases as the model adapts to the congested state, the error level is high when the traffic state transition begins.

Numerical accuracy of the prediction is obviously important in the model comparison. However, comparing different models

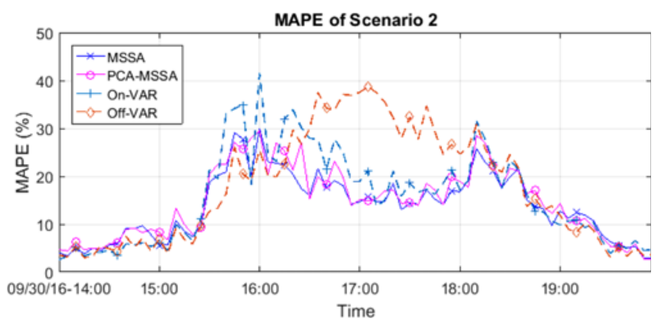


Fig. 7. Prediction Errors During an Incident Event

based solely on the accuracy may be not fair, since other factors such as computation time, required data size, the level of expertise, etc., are important as well (Kirby et a., 1997; Vlahogianni et al., 2014). This is true because the purpose of the proposed method focuses on the near-real-time traffic speed prediction for multiple road segments. Therefore, the computation time to make a one-step prediction with the four methods was compared. The computation time of PCA-MSSA was considerably shorter than those of MSSA and On-VAR model. PCA-MSSA took only 0.05 seconds to predict traffic speed 5 minutes ahead for the 213 different locations; the MSSA algorithm without PCA took 6.78 seconds on average. Although Off-VAR also processed the data quickly, i.e., 0.22 seconds on average, the 0.5-hour training time is not accounted for. In practice, however, the model training time should be considered because periodical updates of parameter values may be needed to retain or enhance the current performance. Combined with the comparison result of prediction accuracy in Scenario 2, the computation efficiency of PCA- MSSA shows that the proposed algorithm is more suitable than the others to predict traffic speed for a large number of road segments in real time. Because of the data dimensionality reduction feature, the proposed method is scalable for a larger road network analysis.

3.5.2 Multi-Step Speed Prediction

The prediction error is accumulated as the number of prediction steps increases. In order to test the prediction performance of PCA-MSSA for the future in longer than five minutes, predictions were made for up to 30 minutes ahead and compared with Off-VAR. Table 3 summarizes the multi-step speed prediction results. Over the multiple prediction steps, the average error of PCA-MSSA showed a moderate increase from 4.98% to 6.88% in Scenario 1. In contrast, the more rapid increase of error from 6.56% to 11.04% was observed in Scenario 2. As a reference, the prediction performance measures of Off-VAR are provided together. As the comparison result for the single-step prediction, the errors of Off-VAR are slightly lower than those of PCA-MSSA in multi-step prediction, whereas the opposite comparison results present in Scenario 2.

It is difficult to directly compare the prediction performance of the proposed algorithm with the results reported in the literature due to different data sources, times, and locations with different study designs. Despite this reason, such comparison may help researchers gain a general sense of the current state in speed prediction studies. The error level of the proposed algorithm is slightly lower or comparable to that of NN-based and time-series models in the literature (Alecsandru and Ishak, 2004; Yang et al., 2004; Guo and Williams, 2010; Djuric et al., 2011; Kamarianakis et al., 2012).

3.5.3 Algorithm Scalability Investigation

To test the scalability of the PCA-MSSA algorithm for speed prediction, NPMRDS data were used in this study. The obtained data cover the entire westbound I-40 segment in Tennessee. The data dimension is (298 × 288) i.e., one day of 5-minute speeds from 298 road links. The majority of the speeds in the dataset represent the free-flow condition except for those of major urban areas during peak hours. Therefore, the computation time is the major interest in this comparison, although the error measures are also presented in Table 4. The comparison result of the computation time is very similar to that in the RTMS case, despite the NPMRDS data dimension being almost 40% larger than the RTMS dataset. PCA-MSSA took 0.36 seconds for one-step prediction, whereas MSSA took 7.24 seconds. The computation time of Off-VAR is also comparably small in the comparison. However, the model estimation time of 2.5 hours is not reflected in the result.

4. Conclusions

Previous short-term traffic prediction studies have investigated a vast number of models and algorithms in the last two decades. Nevertheless, there is still room to progress prediction performance by employing data-driven multivariate models and corresponding large datasets for real-time traffic controls and operations. This paper proposed a short-term traffic speed prediction algorithm to cope efficiently with the complexity and immensity of the prediction process derived from the network size and amount of data. The proposed algorithm, named PCA-MSSA, consists of

Table 3. Prediction Performances for Multi-Step Predictions

			5 min ahead	10 min ahead	15 min ahead	20 min ahead	25 min ahead	30 min ahead
		Prediction steps	1	2	3	4	5	6
Scenario 1 (No incident)	PCA-MSSA	MAE (mph)	2.31	2.57	2.75	2.89	3.01	3.11
		MAPE (%)	4.98	5.57	6.01	6.35	6.64	6.88
	Off-VAR (7)	MAE (mph)	1.96	2.22	2.41	2.55	2.67	2.78
		MAPE (%)	4.26	4.81	5.22	5.55	5.85	6.11
Scenario 2 (No incident)	PCA-MSSA	MAE (mph)	2.46	2.86	3.14	3.37	3.57	3.75
		MAPE (%)	6.56	7.86	8.85	9.66	10.38	11.04
	Off-VAR (7)	MAE (mph)	2.41	2.87	3.24	3.55	3.84	4.10
		MAPE (%)	7.26	9.13	10.76	12.24	13.58	14.81

Table 4. Comparison of 5-minute Prediction Performance for NPMRDS

	PCA-MSSA	MSSA	On-VAR(k)	Off-VAR(k)
Model	$p = 36$, $M = 18$	$p = 36$, $M = 18$	k=3	$k = 3$
MAE (mph)	2.29	2.39	2.24	2.26
MAPE (%)	4.32	4.53	4.16	4.54
Average computation time (s)	0.36	7.24	80.34	0.49

two techniques: (a) principal component analysis (PCA) for data dimensionality reduction and (b) multichannel singular spectral analysis (MSSA) for multivariate time-series data prediction.

The prediction performance of PCA-MSSA was compared to the statistical time-series model, vector autoregressive (VAR). For the incident scenario, PCA-MSSA outperformed VAR and it provided speed predictions in near-real-time. Although the pre-trained VAR model showed slightly lower prediction errors on average for the non-incident scenario, PCA-MSSA still predicted the speed with comparable accuracy levels. This is mainly because PCA-MSSA uses the compressed spatiotemporal traffic data as an input and it has much less parameters to estimate based on input data. In contrast, VAR is a more complex model that requires more data, and it estimates a tremendous number of parameters for multiple road segments analysis. This result shows that PCA-MSSA is suitable for real-time traffic speed prediction and scalable for a large network analysis. To identify the effect of PCA in the proposed algorithm, the results were compared to the case of MSSA without PCA. Interestingly, a trade-off between the accuracy and computation time was reported. Using PCA can reduce computation time significantly with a relatively small compromise in prediction accuracy.

Further research should be directed at the following challenges: (a) improving the prediction accuracy of the proposed algorithm during non-recurring events through cooperation with automatic incident detection algorithms and more advanced PCA methods; (b) adding a self-learning process after the predicted values are validated; (c) developing a dynamic optimization process to select the length of historical data and embedding window length of the algorithm over time; and (d) predicting travel time based on the predicted speed and conducting comparative evaluations. Furthermore, the tested algorithm should be compared at a large road network level with state-of-the-art deep-learning-based methods which are also known as suitable for fitting a nonlinear characteristic of traffic flow in terms of prediction performance, model training cost, and spatial scalability.

Acknowledgments

This paper was prepared based on a part of the first author's dissertation (Bae, 2017) titled "Real-time Traffic Flow Detection and Prediction Algorithm: Data-Driven Analyses on Spatio-Temporal Traffic Dynamics" and based on the authors' poster presentation material titled "Short-term Traffic Speed Prediction for a Large-Scale Road Network" presented at the Transportation Research Board 97th Annual Meeting at Washington D.C. in 2018. The corresponding research work was mostly supported by the Department of Civil and Environmental Engineering, University of Tennessee Knoxville.

ORCID

Bumjoon Bae https://orcid.org/0000-0001-9743-1453
Lee D. Han https://orcid.org/0000-0002-1381-1254

References

Loskutov A, Istomin I, Kotlyarov O (2001) Data analysis: Generalisations of the local approximation method by singular spectrum analysis. ArXiv:nlin/0109022

Akaike H (1973) *Informa*tion Theory and an Extension of the Maximum Likelihood Principle In B. N. Petrov and F. Csaki (Eds.). Paper presented at the Second international symposium on information theory, Budapest: Academiai Kiado

Alecsandru C, Ishak S (2004) Hybrid model-based and memory-based traffic prediction system. *Transportation Research Record: Journal of the Transportation Research Board* 1879:59-70, DOI: 10.3141/1879-08

Alessio SM (2016) Singular spectrum analysis (SSA), 537-571, DOI: 10.1007/978-3-319-25468-5_12

Amin SM, Rodin EY, Liu AP, Rink K, García-Ortiz A (1998) Traffic prediction and management via rbf neural nets and semantic control. *Computer-Aided Civil and Infrastructure Engineering* 13(5):315-327, DOI: 10.1111/0885-9507.00110

Asif MT, Kannan S, Dauwels J, Jaillet P (2013, 16-19 April 2013) Data compression techniques for urban traffic data. Paper presented at the 2013 IEEE Symposium on Computational Intelligence in Vehicles and Transportation Systems (CIVTS)

Bae B (2017) Real-time traffic flow detection and prediction algorithm: Data-driven analyses on spatio-temporal traffic dynamics. PhD Thesis, University of Tennessee Knoxville, USA

Bae B, Kim H, Lim H, Liu Y, Han LD, Freeze PB (2018) Missing data imputation for traffic flow speed using spatio-temporal cokriging. *Transportation Research Part C: Emerging Technologies* 88:124-139, DOI: 10.1016/j.trc.2018.01.015

Chan KY, Dillon T, Chang E, Singh J (2013) Prediction of short-term traffic variables using intelligent swarm-based neural networks. *IEEE Transactions on Control Systems Technology* 21(1):263-274, DOI: 10.1109/TCST.2011.2180386

Chandra S, Al-Deek H (2008) Cross-Correlation Analysis and multivariate prediction of spatial time series of freeway traffic speeds. *Transportation Research Record: Journal of the Transportation Research Board* 2061:64-76, DOI: 10.3141/2061-08

Chandra SR, Al-Deek H (2009) Predictions of freeway traffic speeds and volumes using vector autoregressive models. *Journal of Intelligent Transportation Systems* 13(2):53-72, DOI: 10.1080/15472450902858368

Chen T, Martin E, Montague G (2009) Robust probabilistic PCA with missing data and contribution analysis for outlier detection. *Computational Statistics & Data Analysis* 53(10):3706-3716, DOI: 10.1016/j.csda.2009.03.014

Chhatkuli RB, Demachi K, Miyamoto N, Uesaka M, Haga A (2015) Dynamic image prediction using principal component and multi-channel singular spectral analysis: A feasibility study. *Open Journal of Medical Imaging* 5(3):133-142, DOI: 10.4236/ojmi.2015.53017

Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. New Jersey: Wiley

Dia H (2001) An object-oriented neural network approach to short-term traffic forecasting. *European Journal of Operational Research* 131(2): 253-261, DOI: 10.1016/S0377-2217(00)00125-9

Djuric N, Radosavljevic V, Coric V, Vucetic S (2011) Travel speed forecasting by means of continuous conditional random fields. *Transportation Research Record: Journal of the Transportation Research Board* 2263:131-139, DOI: 10.3141/2263-15

Elsner JB, Tsonis AA (1996) Singular spectrum analysis: A new tool in time series analysis

Fang J, Jin J (2014) Intelligent algorithms for reducing short-term traffic state prediction error in active traffic management. *Journal of Intelligent Transportation Systems* 19(3):304-315, DOI: 10.1080/15472450.2014.893736

Guo J, Williams B (2010) Real-time short-term traffic speed level forecasting and uncertainty quantification using layered kalman filters. *Transportation Research Record: Journal of the Transportation Research Board* 2175:28-37, DOI: 10.3141/2175-04

Guo J, Williams BM, Smith BL (2007) Data collection time intervals for stochastic short-term traffic flow forecasting. *Transportation Research Record: Journal of the Transportation Research Board* 2024(1):18-26, DOI: 10.3141/2024-03

Hassani H, Heravi S, Zhigljavsky A (2013) Forecasting UK industrial production with multivariate singular spectrum analysis. *Journal of Forecasting* 32(5):395-408, DOI: 10.1002/for.2244

Hassani H, Silva ES, Gupta R, Das S (2015) Predicting global temperature anomaly: A definitive investigation using an ensemble of twelve competing forecasting models. Department of economics working paper series 2015-61

Hassani H, Thomakos D (2010) A review on singular spectrum analysis for economic and financial time series. *Statistics and Its Interface* 3(3):377-397, DOI: 10.4310/SII.2010.v3.n3.a11

Heilmann B, El Faouzi NE, de Mouzon O, Hainitz N, Koller H, Bauer D, Antoniou C (2011) Predicting motorway traffic performance by data fusion of local sensor data and electronic toll collection data. *Computer-Aided Civil and Infrastructure Engineering* 26(6):451-463, DOI: 10.1111/j.1467-8667.2010.00696.x

Ishak S, Alecsandru C (2004) Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings. *Journal of Transportation Engineering* 130(4):452-465, DOI: 10.1061/(ASCE)0733-947X(2004)130:4(452)

Kamarianakis Y, Shen W, Wynter L (2012) Real-time road traffic forecasting using regime-switching space-time models and adaptive LASSO. *Applied Stochastic Models in Business and Industry* 28(4):297-315, DOI: 10.1002/asmb.1937

Kirby HR, Watson SM, Dougherty MS (1997) Should we use neural networks or statistical models for short-term motorway traffic forecasting? *International Journal of Forecasting* 13(1):43-50, DOI: 10.1016/S0169-2070(96)00699-1

Lin L, Li J, Chen F, Ye J, Huai J (2018) Road traffic speed prediction: A probabilistic model fusing multi-source data. *IEEE Transactions on Knowledge and Data Engineering* 30(7):1310-1323, DOI: 10.1109/tkde.2017.2718525

Liu Y, Liu Z, Vu HL, Lyu C (2020) A spatio-temporal ensemble method for large-scale traffic state prediction. *Computer-Aided Civil and Infrastructure Engineering* 35(1):26-44, DOI: 10.1111/mice.12459

Lv Y, Duan Y, Kang W, Li Z, Wang FY (2015) Traffic flow prediction with big data: A deep learning approach. *Ieee Transactions on Intelligent Transportation Systems* 16(2):865-873, DOI: 10.1109/TITS.2014.2345663

May AD (1990) Traffic flow fundamentals. New Jersey: Prentice Hall

Michalopoulos PG, Yi P, Lyrintzis AS (1993) Continuum modelling of traffic dynamics for congested freeways. *Transportation Research Part B: Methodological* 27(4):315-332, DOI: 10.1016/0191-2615(93)90041-8

Min W, Wynter L (2011) Real-time road traffic prediction with spatio-temporal correlations. *Transportation Research Part C: Emerging Technologies* 19(4):606-616, DOI: 10.1016/j.trc.2010.10.002

Mizuguchi A, Demachi K, Uesaka M (2010) Establish of the prediction system of chest skin motion with SSA method. *International Journal of Applied Electromagnetics & Mechanics* 33(3/4):1529-1533

Papageorgiou M, Blosseville J-M, Hadj-Salem H (1990) Modelling and real-time control of traffic flow on the southern part of Boulevard Peripherique in Paris: Part I: Modelling. *Transportation Research Part A: General* 24(5):345-359, DOI: 10.1016/0191-2607(90)90047-A

Patterson K, Hassani H, Heravi S, Zhigljavsky A (2011) Multivariate singular spectrum analysis for forecasting revisions to real-time data. *Journal of Applied Statistics* 38(10):2183-2211, DOI: 10.1080/02664763.2010.545371

Payne HJ (1971) Models of freeway traffic and control. In *Mathematical Models of Public Systems* (G. A. Berkeley Ed. ed., Vol. 1). La Jolla, CA.

Polson NG, Sokolov VO (2017) Deep learning for short-term traffic flow prediction. *Transportation Research Part C: Emerging Technologies* 79:1-17, DOI: 10.1016/j.trc.2017.02.024

Quek C, Pasquier M, Lim BBS (2006) POP-TRAFFIC: A novel fuzzy neural approach to road traffic analysis and prediction. *Ieee Transactions on Intelligent Transportation Systems* 7(2):133-146, DOI: 10.1109/TITS.2006.874712

Shang Q, Lin C, Yang Z, Bing Q, Zhou X (2016) A hybrid short-term traffic flow prediction model based on singular spectrum analysis and kernel extreme learning machine. *Plos One*, DOI: 10.1371/journal.pone.0161259.t001

Suksiri P, Chiewchanwattana S, Sunat K (2016, 13-15 July 2016) Application of singular spectrum analysis and kernel-based extreme learning machine for stock price prediction. Paper presented at the 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)

Tan H, Wu Y, Shen B, Jin PJ, Ran B (2016) Short-term traffic prediction based on dynamic tensor completion. *Ieee Transactions on Intelligent Transportation Systems* 17(8):2123-2133, DOI: 10.1109/tits.2015.2513411

Treiber M, Kesting A, Wilson RE (2011) Reconstructing the traffic state by fusion of heterogeneous data. *Computer-Aided Civil and Infrastructure Engineering* 26(6):408-419, DOI: 10.1111/j.1467-8667.2010.00698.x

Turk M, Pentland A (1991) Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1):71-86, DOI: 10.1162/jocn.1991.3.1.71

Vanajakshi L, Rilett LR (2004) A comparison of the performance of artificial neural networks and support vector machines for the prediction of traffic speed. Paper presented at the IEEE Intelligent Vehicles Symposium, June 14-17, 2004

Vitanov NK, Sakai K, Dimitrova ZI (2008) SSA, PCA, TDPSC, ACFA: Useful combination of methods for analysis of short and nonstationary time series. *Chaos, Solitons & Fractals* 37(1):187-202, DOI: 10.1016/j.chaos.2006.08.043

Vlahogianni EI, Karlaftis MG, Golias JC (2014) Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies* 43:3-19, DOI: 10.1016/j.trc.2014.01.005

Wang J, Shi Q (2013) Short-term traffic speed forecasting hybrid model based on Chaos–Wavelet Analysis-Support Vector Machine theory. *Transportation Research Part C: Emerging Technologies* 27:219-232, DOI: 10.1016/j.trc.2012.08.004

Yang F, Yin Z, Liu H, Ran B (2004) Online recursive algorithm for short-term traffic prediction. *Transportation Research Record: Journal of the Transportation Research Board* 1879:1-8, DOI: 10.3141/1879-01

Ye Q, Szeto WY, Wong SC (2012) Short-term traffic speed forecasting based on data recorded at irregular intervals. *Ieee Transactions on Intelligent Transportation Systems* 13(4):1727-1737, DOI: 10.1109/TITS.2012.2203122