Trend Modeling for Traffic Time Series Analysis: An Integrated Study

Li Li, *Senior Member, IEEE*, Xiaonan Su, Yi Zhang, *Member, IEEE*, Yuetong Lin, *Senior Member, IEEE*, and Zhiheng Li, *Member, IEEE*

***Abstract*—This paper discusses the trend modeling for traffic time series. First, we recount two types of definitions for a long-term trend that appeared in previous studies and illustrate their intrinsic differences. We show that, by assuming an implicit temporal connection among the time series observed at different days/locations, the PCA trend brings several advantages to traffic time series analysis. We also describe and define the so-called short-term trend that cannot be characterized by existing definitions. Second, we sequentially review the role that trend modeling plays in four major problems in traffic time series analysis: abnormal data detection, data compression, missing data imputation, and traffic prediction. The relations between these problems are revealed, and the benefit of detrending is explained. For the first three problems, we summarize our findings in the last ten years and try to provide an integrated framework for future study. For traffic prediction problem, we present a new explanation on why prediction accuracy can be improved at data points representing the short-term trends if the traffic information from multiple sensors can be appropriately used. This finding indicates that the trend modeling is not only a technique to specify the temporal pattern but is also related to the spatial relation of traffic time series.**

***Index Terms*—Traffic time series, trend, detrending, traffic prediction.**

I. Introduction

TRAFFIC time series analysis is a key function of most Intelligent Transportation Systems (ITS) research and applications [1]. Four major issues in this field have received continuous interests during the last three decades.

The first problem is abnormal traffic data detection, which aims to detect abrupt changes in traffic conditions caused by traffic accidents, or to isolate the anomaly caused by failures of detectors [2]–[6].

The second problem is traffic data compression, which aims to provide a low-dimensional approximation of the raw traffic dataset to address the storage and searching of large volumes of traffic data [7]–[13].

The third problem is missing data imputation, which aims to estimate the missing traffic data for further analysis [2], [14]–[21].

The fourth topic is traffic prediction, which aims to foresee the forthcoming traffic conditions based on historical records and up-to-date observations of traffic conditions [1], [22]–[27].

A great number of models have been proposed for each problem. However, to the best of our knowledge, these four problems were primarily investigated separately and few studies have discussed their intrinsic relations.

As pointed out in [1], the success of these analysis all relies on the appropriate modeling of temporal and spatial characteristics of traffic time series. In agreement with this conclusion, we will show in this paper that all these problems are related to the representation and decomposition of traffic time series. More precisely, each problem can be characterized by just one or two components of traffic time series. Therefore, to consider unrelated components will not improve but often degrade the analysis performance.

Our purposes are as follows:

First, we present an integrated framework for all four problems and their relations to trend modeling of traffic time series.

Second, we review three of the most commonly used long-term trend definitions: simple average trend, principal component analysis (PCA) trend and wavelet trend. We compare their intrinsic differences that have not been well explained in previous studies. Short-term trend, which cannot be characterized by previous studies, is also defined.

And finally, we explain why networked information can improve traffic prediction accuracy and its relation with trend modeling.[1] This interesting finding indicates that the trend modeling is not only a technique to specify the temporal pattern but is also related to the spatial relation of traffic time series.

To present our findings, the rest of this paper is arranged as follows. Section II first reviews two kinds of definitions for

Manuscript received October 23, 2014; revised April 2, 2015 and June 14, 2015; accepted July 9, 2015. Date of publication August 5, 2015; date of current version November 23, 2015. This work was supported in part by the National Science and Technology Support Program under Grant 2013BAG18B00, by the National Natural Science Foundation of China under Grant 51278280, and by Tsinghua University under Project 20131089307. The Associate Editor for this paper was B. De Schutter. *(Corresponding author: Zhiheng Li.)*

L. Li and Y. Zhang are with the Department of Automation, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China, and also with Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Nanjing 210096, China.

X. Su is with the Department of Automation, Tsinghua University, Beijing 518055, China.

Y. Lin is with the Department of Electronics and Computer Engineering Technology, Indiana State University, Terra Haute, IN 47809-9989 USA.

Z. Li is with the Department of Automation, Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China, and also with the Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China (e-mail: zhhli@mail.tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2015.2457240

[1]In this paper, we assume that multiple traffic flow time series can be obtained at different sensors (at different locations). So, we use the term "sensor" to distinguish the available time series. The time series observed at all these sensors formulate a "networked" data source, where the dependences between time series are important to data compression and traffic prediction. So, we call such data as "networked" information.

long-term trend existed in previous studies and account for their intrinsic differences. We then describe the short-term trend that cannot be characterized by previous definitions. Sections III–V sequentially reviews the role of trend modeling in four key issues in traffic time series analysis: namely abnormal data detection, data compression, missing data imputation and traffic prediction. The relations between these problems are revealed and the benefits of detrending are explained. Finally, Section VI concludes the paper.

II Trends in Traffic Time Series

A. The Long-Term Trend

The first type of long term trend is defined as the daily similarity of traffic time series.

A prominent feature of traffic observed at a certain location is that the daily flow time series have similar M-shaped intra-day trends over consecutive days, in which the morning and evening rush hours correspond to the two peaks of the M-shape. Such a daily trend (also called daily profile in many reports e.g., [28]) constitutes the major part of traffic time series.

Based on this observation, many approaches decompose traffic time series into the so called daily trend time series and the residual times series (also called deviations to trend). However, there exist different ways to mathematically define the trend.

The easiest way to describe the trend is to use the average of the daily flow time series collected in a specified time window. This leads to the definition of the *simple average trend*.

Suppose the traffic time series sampled in N consecutive work days can be written as a series of one-dimensional vectors

$$\vec{y}_1 = [y_{1,1}\ y_{2,1}, \ldots, y_{n,1}]^T, \ldots, \vec{y}_N = [y_{1,N}\ y_{2,N}, \ldots, y_{n,N}]^T \tag{1}$$

where we assume there are n sample data points per day. If the sample time interval is 5 minutes, we then have $n = 288$. $y_{i,j}$ with $i = 1, \ldots, n$, represents the traffic flow of the ith sample data point on the jth day.

The simple average daily trend over the last N working days can then be calculated as

$$\vec{y}_{Average} = \left[\frac{1}{N}\sum_{j=1}^{N} y_{1,j}, \ldots, \frac{1}{N}\sum_{j=1}^{N} y_{n,j}\right] \tag{2}$$

while the residual time series is then calculated as $\vec{y}_i - \vec{y}_{Avergae}, i = 1, \ldots, N$.

As an example, consider a typical simple average daily trend and its original flow volume time series shown in Fig. 1. The relatively small deviation indicates that the simple average daily trend has captured the major variation pattern of traffic time series.

The simple average daily trend provides a rough sketch of the daily traffic time series over a continuous period. However, it neglects the local variation patterns within each day, therefore cannot be used for traffic data compression and missing data imputation.

Another frequently used definition of daily similarity trend is based on principal component analysis (PCA) [31]). As a well-known mathematical dimensionality reduction and feature extraction procedure, PCA projects the original data vectors $\vec{y}_j$ onto an n-dimensional orthogonal linear space such that the data with the kth largest variance by projection lies on the kth dimension.

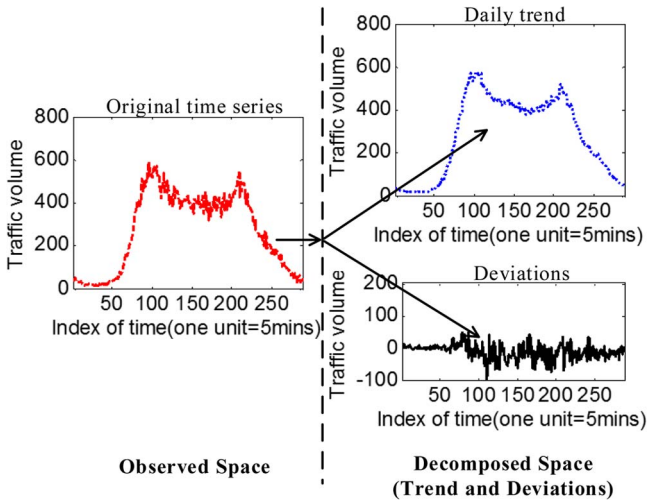


Fig. 1. An illustration of the simple average trend. The data used in this paper are all from the public accessible PeMS dataset [29]. The specific data shown in this figure are gathered by the sensor (ID 402580) located in the fourth district of PeMS dataset and reflect the traffic flow on August 1, 2011. The daily trend is calculated based on traffic volume data of 15 consecutive workdays since August 1, 2011.

More precisely, for each daily data vector $\vec{y}_j$, there exists a vector (latent variable) $\vec{x}_j \in R^n$ that satisfies

$$\vec{y}_j = W\,\vec{x}_j + \vec{\mu} \tag{3}$$

where $\vec{\mu} \in R^{n\times 1}$ allows the mean shift and $W \in R^{n\times n}$ is the projection matrix. We will not introduce the details on how to calculate the projection matrix W and the latent variable $\vec{x}_j$. Interested readers are referred to a number of good surveys, e.g., [30], [31] for further information.

If each $\vec{y}_j$ follows an independent identically distributed (i.i.d) Gaussian distribution, the sum of the variances is usually dominated by a few principal components. This nice property indicates that the major varying patterns of each $\vec{y}_j$ can be recovered from the weighted sum of a finite number of $\vec{x}_j$.

After choosing h, the number of latent variables, we can distinguish the trend and the deviations via PCA. The time series reconstructed from the first h latent variables will be treated as the daily trend. The remaining latent variables will be used to recover the residual time series. The PCA trend describes a more detailed sketch of daily traffic, and it allows us to define different trends for different days; see Fig. 2 for an illustration.

More importantly, the PCA trend establishes a link between traffic time series collected in different days/locations because the observed daily data vectors share the same set of latent variables. This feature implies that the variations of daily traffic time series have some patterns and are not fully random. Tests in [16], [17], [19], [20] have shown that this advantage makes PCA based methods outperform many other approaches in traffic time series analysis.

We would like to explain the term "dependence" and "joint occurrence probability" used here and onwards. In general, if

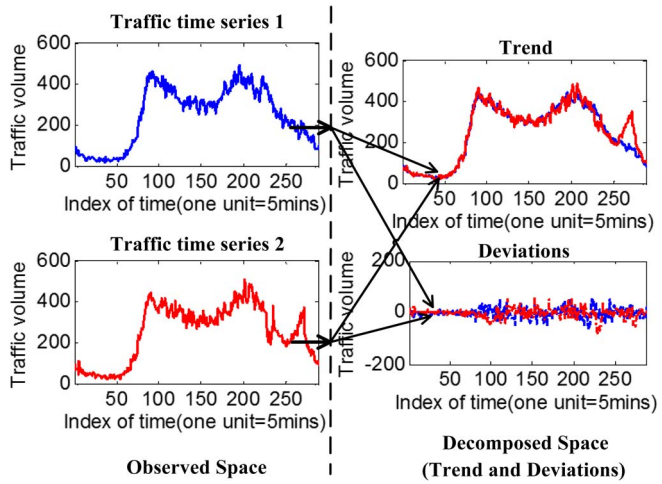


Fig. 2. An illustration of PCA trend. The data, reflecting traffic flow on August 3, 2011 (Traffic time series 1) and August 4, 2011 (Traffic time series 2), are collected by the sensor (ID 400116) located in the fourth district of PeMS dataset.

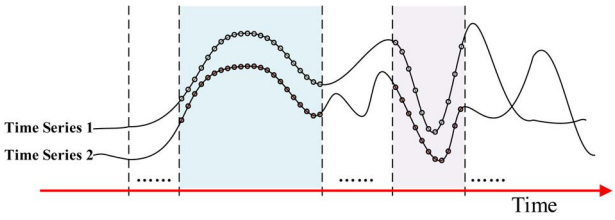


Fig. 3. An illustration of the "dependence" and "joint occurrence probability" between daily traffic time series. Since Time Series 1 and Time Series 2 have similar varying tendencies, we assume there exists "dependence" between Time Series 1 and Time Series 2.

we take the values of different time series collected at the same time point, the "dependence" refers to a special joint occurrence relation. We say two time series are "dependent" if for example, the value of **Time Series 1** is large (or small) now, and the value of **Time Series 2** has a large probability to be large (or small), too; see Fig. 3 for an illustration. We do not use the more rigorous terms "cause" and "effect" here because we often cannot determine which time series is the "cause" and which time series is the "effect". We do not use the term "correlated" in statistics either because the relation between two time series might be both deterministic and stochastic.

The second type of long-term trend is based on evolution tendency of traffic time series.

The basic ideas of this type of trend are a) the whole evolution tendency is smooth (without high-frequency fluctuations); b) tendency at the ith sample point is strongly related with the traffic data collected in its vicinity. That is, we have the trend defined as

$$t_{i,j} = f\left(y_{i-n1,j}, \ldots, y_{i-1,j}, y_{i,j}, y_{i+1,j}, \ldots, y_{i+n2,j}\right) \quad (4)$$

where $t_{i,j}$ is the trend (tendency) value at the ith sample time point on the jth day, $f(x)$ is a certain smooth mapping function, $n1$ and $n2$ indicate the length of the smoothing window.

Discrete wavelet transform (DWT) is a typical method that defines trend via tendency. Using low/high-pass filtering and down-sampling, DWT tries to reveal the evolution relationship between the consecutive sample points. As a result, the original time series are casted into two types of coefficients: scaling coefficients and wavelet coefficients [32]. The scaling coefficients reflect the low-frequency part (or approximations) of time series, and the wavelet coefficient indicate the high-frequency part (or details) of time series. The casting procedure can be formulated as:

$$y^{a,le}_{i1,j} = \sum_k y^{a,le-1}_{k,j} g^{low}(ds \cdot i1 - k) \quad (5)$$

$$y^{d,le}_{i1,j} = \sum_k y^{a,le-1}_{k,j} g^{high}(ds \cdot i1 - k) \quad (6)$$

where $y^{a,le}_{i1,j}/y^{d,le}_{i1,j}$ is the scaling/wavelet coefficient corresponding to $\vec{y}_j$ in level le, $le \geq 1$. $y^{a,0}_{k,j} = y_{k,j}$, $g^{low/high}(x)$ denotes the selected low/high pass filter function, ds indicates the factor of down-sampling. The superscript "a" stands for approximation coefficients and the superscript "d" stands for detail coefficients [32].

Although Eq. (5), (6) just give out the general form of wavelet based trend modeling, they had highlighted one main feature of such approaches: the trend at a data point is defined within the traffic data collected in its vicinity. This kind of approach is quite different from similarly based trending modeling which check the dependence around the traffic data collected in different days.

Given the selected level of details, the wavelet trend can be obtained by reconstructing the time series with only the preserved scaling coefficients and omitting the high-frequency fluctuations. In other words, the wavelet trend can be written as

$$t^{le}_{i,j} = r^{low}\left(\vec{y}^{a,le}_j\right) \quad (7)$$

where $r^{low}(x)$ is the reconstruction function, $\vec{y}^{a,le}_j$ is a vector consisting of $y^{a,le}_{i1,j}$.

We can see that the trend defined by wavelet transform mainly captures the self-dynamics within the studied time series; while the simple average trend and PCA trend aim to recover the common dynamics between multiple daily traffic time series. In other words, wavelet trend mainly defines the local (within a relatively small time frame, e.g., a half day) varying feature (tendency); while the simple average trend and PCA trend compare the daily traffic time series collected in several consecutive days and define their global (across several days) varying feature (similarity). This inherent difference makes wavelet based detrending [33]–[37] behave differently from simple average or PCA detrending.

B. The Short Term Trend

It has been pointed out in many literatures, e.g., [9], [27], [38]–[40], that because of the existence of several "short-term" trends, the residual time series do not strictly follow i.i.d. Gaussian distribution; see Fig. 4 for an example.

Suppose the standard deviation σ, has been obtained for the entire residual time series, we define the short-term trend as a consecutive sequence of data in residual time series that are all bigger than 2σ, or all smaller than -2σ. The term short-term means the length of these data sequences is much shorter than

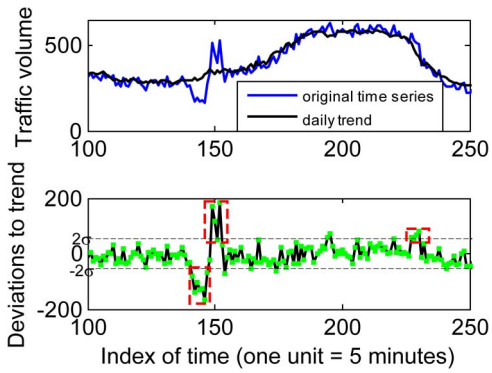


Fig. 4. An illustration of the short-term trend in residual time series. The data show the traffic flow time series collected at the sensor (ID 400197) located in the fourth district of PeMS dataset on August 22, 2011.

that of the daily traffic time series, therefore short-term trends will not significantly affect the simple average trend.

As the name suggests, short-term trends are only temporary. Nevertheless, they represent the sudden volatility in traffic flow. Many literatures have noticed the existence of short-term trends. However, few studies have focused on their relations with traffic time series analysis, an important subject that we will elaborate in Section V-B.

III. Abnormal Data Detection and Trend Modeling

Abnormal data detection usually refers to finding the data that do not yield expected behaviors [41]. In traffic applications, abnormal data detection concerns about automatically isolating unusual traffic phenomena from massive datasets and helping determine the causes for the irregularities.

Abnormal data can be generated by traffic accidents, special events or loop detector faults. In the first two cases, the anomaly is usually non-recurrent; while loop detector faults usually lead to daily recurrent anomaly. Online detection of unusual traffic flow patterns can enable the system managers to quickly respond to changing conditions. Furthermore, automatically detecting recurrent anomaly can greatly reduce the workload of system operators [3]. In this paper, we focus on the latter problem.

As it turns out, abnormal data detection is a relatively easy problem because the difference between two traffic time series is dominated by the difference between their daily trends [42], [43], and the time series recorded from normal and abnormal sensors often have quite different trends.

Therefore, we can directly collect the time series of all the sensors over several days and group them into different clusters using unsupervised learning. The outliers generated by clustering may be produced by the potential abnormal sensors. This method does not need any advanced assumptions for traffic time series to start detection and is thus easy to use.

Fig. 5 below gives an example in which we use dendrogram to automatically distinguish the outliers. A dendrogram is a tree-structured graph that visualizes the result of a hierarchical clustering. In dendrogram, the top row of nodes represents individual observations, and the remaining nodes represent the clusters to which these observations belong.

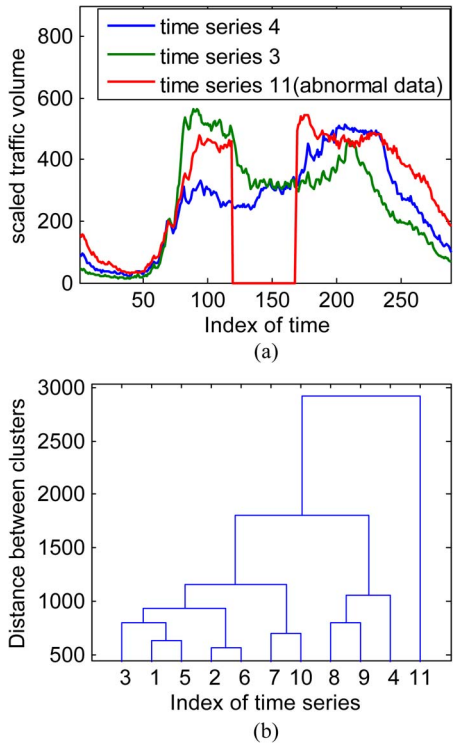


Fig. 5. (a) The simple average daily trends observed at 11 sensors; (b) the corresponding dendrogram, from which the outlier (the 11th sensor) can be easily distinguished. The studied sensors (ID 400409, 400411, 400414, 400417, 400418, 400422, 400425, 400428, 400429, 400432, 400367) are located in the fourth district of PeMS dataset. The daily trends are calculated based on the traffic volume of 15 consecutive workdays, from August 1, 2011 to August 9, 2011. The obtained daily trends are rescaled to preserve their means. We simulate a long sequence of missing data for the time series obtained from the 11th sensor.

We use the Euclidean distance measure

$$D_{\alpha,\beta} = \sqrt{\sum_{i=1}^{n} \left(y^{\alpha}_{i,average} - y^{\beta}_{i,average} \right)^2} \tag{8}$$

to depict the distance (dissimilarity) between the αth and βth observation. $y^{\alpha}_{i,average}$ and $y^{\beta}_{i,average}$ denote the ith data point in the average daily trend of the αth and βth observation separately. n is the number of sample points in each observation. The distance between two clusters is the average distance between components in separate clusters. That is

$$D_{cluster} = \frac{1}{mk} \sum_{\alpha=1}^{m} \sum_{\beta=1}^{k} D_{\alpha,\beta} \tag{9}$$

where m and k represent the number of components in the first and second cluster, respectively.

In Fig. 5, we can see that the abnormal data from the 11th sensor, which contains a long sequence of missing data, can be easily identified due to their significantly larger separations from others.

PCA based method provides another way to separate the normal and abnormal data by checking the distances between their projections in the latent space. Constrained by the length limit, we only discuss the working mechanism of robust PCA method.

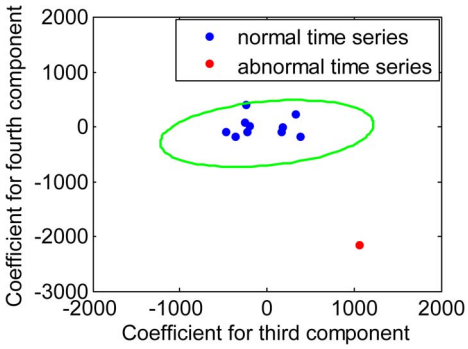


Fig. 6. An illustration of the projection of the normal and abnormal data on the two-dimensional space spanned by the third and fourth latent variables. The space spanned by the third and fourth latent variables is chosen because the normal and abnormal data have similar projections on the two-dimensional space spanned by the first two latent variables. We use the data appeared in Fig. 5.

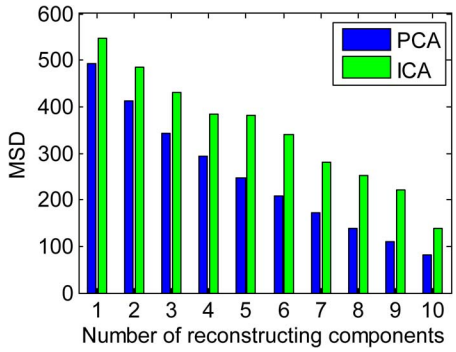


Fig. 7. A comparison of PCA/ICA-based traffic flow data compression. The specific data used in this figure are gathered by the sensor (ID 402580) located in the fourth district of PeMS dataset and reflect traffic volume of 15 consecutive workdays beginning on August 1, 2011.

Robust PCA method tries to remove the influence of outliers in order to correctly recover principal components. Fig. 6 gives an example for the data shown in Fig. 5(a). The projection matrix is determined by the normal data found by robust PCA. As can be seen from the figure, the projection (plotted in red) of the abnormal data is located far from the projections (plotted in blue) of the normal data in the latent space.

Comparing the data projections sometimes is more reliable than comparing their simple average trends due to the following two reasons:

First, robust PCA projection method decomposes the data into more components. This usually has an effect of trivializing the common features of traffic time series but highlighting their unique features. This strategy generally reduces the false alarm rate.

Second, robust PCA projection comparison can set up a more accurate quantitative rule to differentiate normal and abnormal data based on the assumption of multivariate Gaussian distributions [44], [45]. Interested readers are referred to [3], [17] for detailed discussions.

We need to point out that our main objective is to give a simple illustration of the role that trend modeling plays for abnormal data detection. We do not claim that the aforementioned methods are better than some state-of-the-art methodologies.

IV. Traffic Data Compression, Missing Traffic Data Imputation and Trend Modeling

A. Traffic Data Compression

PCA method also provides a simple yet convenient way to compress traffic data. If, instead of the original data matrix $Y = [\vec{y}_1\ , \vec{y}_2, \ldots, \vec{y}_N]$, we manage to store only a part of the projection matrix W and the most important principal components $\vec{x}_j$, $j = 1, \ldots, h$, we can save $(nN - nh - hN - n)$ units of memory space. More importantly, we can achieve data compression when $h \ll n$, and recover Y if needed.

Reference [46] has compared PCA with its lesser known counterpart, Independent Component Analysis (ICA) [47]. Both PCA and ICA aim to express a set of observations as the projection from a set of latent variables with lower dimensions. PCA assumes that the latent variables follow uncorrelated Gaussian distributions, whereas ICA assumes the latent variables are independent and non-Gaussian. Fig. 7 shows the reconstruction errors with respect to the number of PCA and ICA components. Here, we take the frequently used Mean Square Deviation (MSD) defined in Eq. (10) as the evaluation criterion

$$MSD = \frac{1}{N \times n} \sum_{j=1}^{N} \sum_{i=1}^{n} \left(y_{i,j}^{observed} - y_{i,j}^{reconstructed} \right)^2 \quad (10)$$

where $y_{i,j}^{observed}$ and $y_{i,j}^{reconstructed}$ are the observed value and reconstructed value of the ith element in the jth testing vector, respectively. N denotes the number of vectors under test.

As illustrated in [9]–[12], the PCA method is an efficient tool for traffic data compression. Fig. 7 shows that PCA based model outperforms ICA based model because PCA method's assumption, which is the residual time series roughly follow a multi-variable Gaussian distribution in all the time aggregation levels, fits better with empirical data.

B. Missing Traffic Data Imputation

To apply the PCA method for missing traffic data imputation, the key is to recover the projection matrix M and latent variable $\vec{x}_j$ based on the known part of $\vec{y}_j$. The missing part of $\vec{y}_j$ is then estimated using the projection relation $\vec{y}_j \approx M\ \vec{x}_j + \vec{\mu}$.

To this end, reference [48] proposes a statistical extension of the classical PCA, which is called probabilistic principal component analysis (PPCA). It is further assumed in [48] that the latent variable and disturbance are all Gaussian distributed and contribute to the observation in the following form

$$\vec{y}_j = M\ \vec{x}_j + \vec{\mu} + \vec{\varepsilon} \quad (11)$$

where $\vec{y}_j \in R^{n \times 1}$ indicates the observed data, $\vec{x}_j \in R^{h \times 1}$ represents the projection vector of $\vec{y}_j$, and $\vec{\varepsilon}$ is the isotropic noise.

It has been proved in [48] that when the data is complete, the PPCA model (11) is equivalent to the classical PCA dimension reduction model (3).

Assuming $\vec{x} \sim N(0, I)$, $\vec{\varepsilon} \sim N(0, \sigma^2 I)$, the marginal distribution of $\vec{y}_j$ can be described as $\vec{y}_j \sim N(\vec{\mu}, C)$ with

$C = MM^T + \sigma^2 I$, and the log-likelihood estimate of $\vec{y}_j$, $j = 1, \ldots, N$, can be calculated as:

$$L_C = \sum_{i=1}^{N} \ln \left\{ p\left(\vec{x}_j | \vec{y}_j, M, \vec{\mu}, \sigma^2\right) \right\} \tag{12}$$

where $\vec{x}_j$ is the corresponding projection vector of $\vec{y}_j$, σ^2 denotes the variance of noise $\vec{\varepsilon}$.

However, the weakness of PPCA method is that we cannot directly estimate the mapping matrix M when the data is incomplete. A frequently used alternative is to use the expectation maximization (EM) algorithm whose details can be found in [17], [48].

The key for PPCA based missing data imputation is to turn the missing data estimation into a maximum likelihood estimation problem, in which the best estimator of the missing data gives the largest likelihood. This feature simplifies both the evaluation rule and algorithm for solving missing data imputation. Tests in [16], [17], [19], [20] have shown that PPCA approach outperforms many other missing data imputation approaches in both accuracy and speed.

When researching PCA for compression and PPCA for missing data imputation, we find that their strengths are due to the assumption on the joint occurrence probability of the traffic time series collected on different days. This leads us to conclude that there exit implicit connections between traffic time series observed in the network. Readers who are interested in this subject can find more detailed discussions in [17], [20].

V. Traffic Prediction and Trend Modeling

We discuss two interlaced problems that are related to both trend modeling and traffic prediction:

1) Should we incorporate the complete original traffic time series or just the residual time series (with the simple average daily trend removed) into prediction models?
2) Why the prediction performance can be significantly improved when the information from multiple sensors is applied?

A. The Benefit of Detrending

In general, traffic prediction aims to determine a relationship (mapping function) between the past data and the future data so that we can use the known data to estimate the unknown data. There exist a lot of prediction models such as the Auto- Regressive Moving Average (ARMA) model, Feed-Forward Neural Network (FFNN) model etc. Great efforts have been dedicated to selecting appropriate model structures and parameters in the last two decades [1], [24]–[26].

In Reference [27], a new strategy is proposed to improve prediction accuracy: First, the simple average daily trend is removed prior to prediction. Then, the residual series is fed into the prediction model for training and prediction. Finally, the simple average trend is superimposed on the predicated values. As shown in [27], we may significantly improve the prediction performance by using this detrending strategy under the circumstance that only one sensor is available.

In this paper, we further investigate whether the detrending strategy still works when the information of multiple sensors is applied.

Constrained by the length limit, we only examine two frequently used prediction models [1], [24]–[26]: multi-variable linear regression (MVLR) model and Feed-Forward Neural Network (FFNN) model.

The MVLR model sets a linear dependence between the past data (including past data of studied time series and the causally dependent time series) and the future data as

$$y_{i,j} = \sum_{l=1}^{L} a_l y_{i-l,j} + \sum_{k=1}^{num} \sum_{l=1}^{L} b_{l,k} x_{i-l,j}^{k} + \varepsilon_{i,j} \tag{13}$$

where $y_{i-l,j}$ represents the past data of the studied time series recorded at time $i-l$, $x_{i-l,j}^{k}$ is the past data of the kth reference time series recorded at time $i-l$. *num* denotes the number of the reference time series, L shows the max time lag, a_l and $b_{l,k}$ are the regression coefficients of the model, $\varepsilon_{i,j}$ denotes the modeling error. Clearly, if we do not consider the causally dependent time series in (13), the MVLR model degenerates to the classical Moving Average (MA) model.

The multi-variable FFNN model assumes a nonlinear dependence between the past data and the future data as

$$y_{i,j} = \sum_{hi=1}^{H} c_{hi} \psi \left(\sum_{l=1}^{L} w_{l,hi}^{0} y_{i-l,j} + \sum_{k=1}^{num} \sum_{l=1}^{L} w_{l,hi}^{k} x_{i-l,j}^{k} + d_{hi} \right) + d_0 + \widetilde{\varepsilon}_{i,j} \tag{14}$$

where H is the number of hidden neurons, $\psi(x)$ is the activation function for the hidden neurons. $w_{l,hi}^{k}$ is the weighting coefficient from the lth input neuron of $x_{i,j}^{k}$ to the *hi* th hidden neuron. c_{hi} is the weighting coefficient from the *hi* th hidden neuron to the output neuron. d_0 and d_{hi} are constant biases. $\widetilde{\varepsilon}_{i,j}$ denotes the modeling error.

In summary, all prediction models intend to set up smooth mappings between the historical data observed at different sensors and the future data that will be observed at the studied sensor. We will not present how to select the related sensors and how to train prediction models. Interested readers could find relevant results in [27], [49].

Fig. 8 compares the performances of prediction models in (13) and (14) with and without detrending, when data from multiple sensors are employed. The sample time interval is 5 minutes. Here and onwards, we use the Mean Square Deviation (MSD) defined below as the performance index

$$MSD = \frac{1}{\widetilde{N} \times n} \sum_{j=1}^{\widetilde{N}} \sum_{i=1}^{n} \left(y_{i,j}^{observed} - y_{i,j}^{predicted} \right)^2 \tag{15}$$

where $y_{i,j}^{observed}$ and $y_{i,j}^{predicted}$ are the observed value and predicted value of the ith element of the jth testing vector. $\widetilde{N}$ denotes the number of vectors under test.

As shown in Fig. 8, the prediction accuracy is significantly improved for both MA and FFNN models when residual time series instead of original time series, regardless of the number

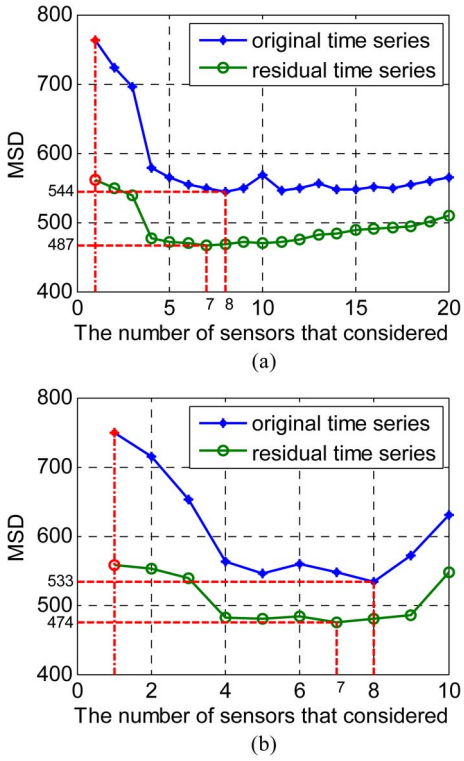


Fig. 8. (a) A comparison between using original time series and using only residual time series in MVLR model. (b) A comparison between using original time series and only residual time series in FFNN model. We choose the maximal lag $L = 10$, which means the maximal time lag is 50 minutes. For FFNN model, the number of hidden neurons is chosen to be $H = 5 + num$. When the number of sensors is one, only one sensor is available (that is, we only consider the self-dependence of the studied time series).

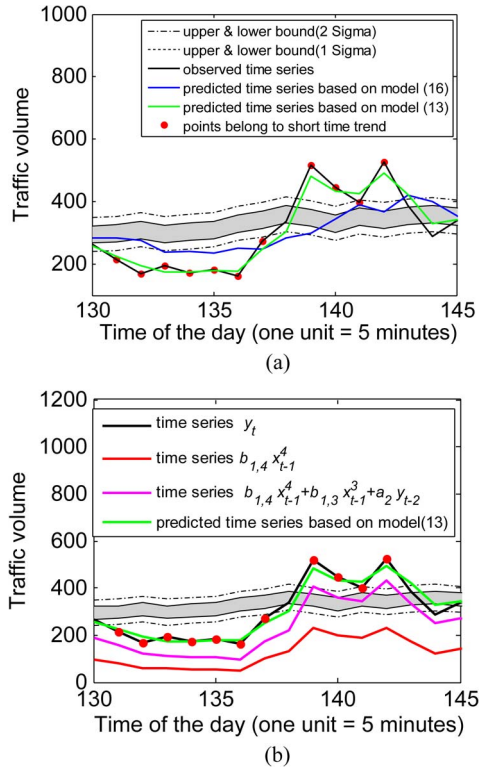


Fig. 9. (a) A comparison between prediction results by using single sensor and multiple sensors. (b) an illustration of the accumulative effect of several related time series. Here, we choose the data from the sensor (ID 400197) in PeMS to illuminate the influence of short-term trends on traffic prediction. Traffic volume data from August 1, 2011 to August 21, 2011 are used as training data, traffic volume data of August 22, 2011 are used as testing data to evaluate prediction performance.

of sensors considered. We have also tested other prediction models, for instance, support vector regressions [27]. The conclusions are similar.

We now offer the explanation for the benefit of detrending that is left out in [27]. Based on the above discussions, we can see that a traffic time series consists of several components. The simple average trend, which is the largest component, often reflects the time-invariant part of traffic time series; and the remaining components as a whole reflect the time-variant fluctuation of traffic time series. Consequently, feeding the original series with the time-invariant trend into prediction models, which are built to capture the dynamics of traffic time series, may bias the training results and lead to false predictions. Therefore, detrending helps to improve prediction accuracy significantly.

It must be pointed out that we can use different detrending methods to retrieve the trend. Generally speaking, the simple average trend has the least influence on the prediction models. The PCA trend and wavelet trend have added certain dependence assumptions between the data. It has been reported in some literatures that a combination of wavelet detrending and prediction models may improve the prediction accuracy. However, we should be careful when mixing PCA/wavelet detrending and prediction models together because the prediction models will introduce dependence assumptions themselves. Whether the assumptions made by detrending method agrees with those made by prediction model often requires further discussions.

B. The Benefit of Multiple Sensors

Fig. 8 also indicates that the prediction accuracy can be noticeably improved when multiple sensors are considered. Although there already exist several explanations for such an improvement, we attempt a new angle focusing on the influence of short-term trends.

When the information of only one sensor is used, many prediction models, due to their assumptions of smooth mapping between historical and future data, cannot adequately detect and track the short-term trends defined in Section II-B, [27]. As shown in Fig. 9(a), the predicted values mainly fall within the intervals bounded by the upper and lower envelopes, which are two standard deviations away from the simple average trend line.

Let us use the following example to illustrate the impact of smoothness assumption for the MVLR model. The predicted value of an MVLR model using the information of a single sensor relies on a few past data. Based on the training data of the sensor (ID 400197), we have obtained a MVLR model for prediction as

$$\begin{aligned} y_{i,j} = {} & 0.292y_{i-1,j} + 0.170y_{i-2,j} + 0.076y_{i-3,j} \\ & + 0.051y_{i-4,j} + 0.019y_{i-5,j} + 0.047y_{i-6,j} \\ & + 0.003y_{i-7,j} + 0.028y_{i-8,j} + 0.021y_{i-9,j} \\ & + 0.030y_{i-10,j} + \varepsilon_{i,j} \end{aligned} \tag{16}$$

We can see that all the estimated regression coefficients are relatively small, which is consistent with the general pattern of MVLR model according to its presumption. If the data recorded before time t are all located within the region bound by envelopes that are two standard deviations away from the simple average trend line, the subsequent predicated values will grow slowly due to the trivial regression coefficients; see Fig. 9(a) for an example. Thus, if short-term trends occur immediately following time t, the MVLR model is unable of catching up with the sudden volatility.

However, this problem can be partially resolved if multiple sensors are used. As we have shown in Section II-A, because of the influence of joint occurrence probability of the traffic time series, if the value of one residual time series collected at one sensor begins to increase above twice the standard deviation threshold, the values of residual time series collected at other related sensors will most likely increase significantly.

According to Eq. (13), a forecasted value of the networked prediction models is the weighted sum of the past values of multiple related residual time series. Although $b_{l,k}x^k_{i-l,j}$, the contribution of one sensor to the predicted value $y_{i,j}$, might be limited, the final sum $\sum_l b_{l,k}x^k_{i-l,j}$ could be large enough when the contribution of multiple sensors are accumulated; see Fig. 9(b) for an example. This mechanism enables the predicated time series to catch up with short-term trends.

To illustrate why we should emphasize the prediction performance of the points belonging to short-term trends, we consider a separate MSD index:

$$MSD = \frac{1}{\widetilde{N} \times n} \sum_{j=1}^{\widetilde{N}} \sum_{i=1}^{n1} \left(y^{observed}_{i,j,trend} - y^{predicted}_{i,j,trend} \right)^2 + \frac{1}{\widetilde{N} \times n} \sum_{j=1}^{\widetilde{N}} \sum_{i=1}^{n2} \left(y^{observed}_{i,j,other} - y^{predicted}_{i,j,other} \right)^2 \quad (17)$$

where $y^{observed}_{i,j,trend}$ denotes the collected traffic time series data that belong to short-term trends, $y^{observed}_{i,j,other}$ refers to the rest of time series data points. $n1$ and $n2$ indicate the length of short-term trends and the number of remaining data points in the series, respectively.

Fig. 10(a) plots the ratios of the short-term trend points versus the rest points. In total, we have 288 points in the test dataset and 24 points belong to the short-term trends. Although the short-term trends only account for 8% of the data points, they contribute to 51% MSD prediction errors; see Fig. 10(b).

Fig. 10(b) reveals that using multiple sensors greatly reduces the prediction errors for the short-term trends in MVLR models. We can see that the prediction errors are also reduced for other points if multiple sensors are employed, however it is at the short-term points that the information from multiple sensors causes the most significant prediction improvement.

It must be pointed out that the above conclusions also hold when thousands of sensors are considered. However, a detailed explanation on sensor selection requires a dedicated paper. Interested readers are referred to [49], [50] for help.

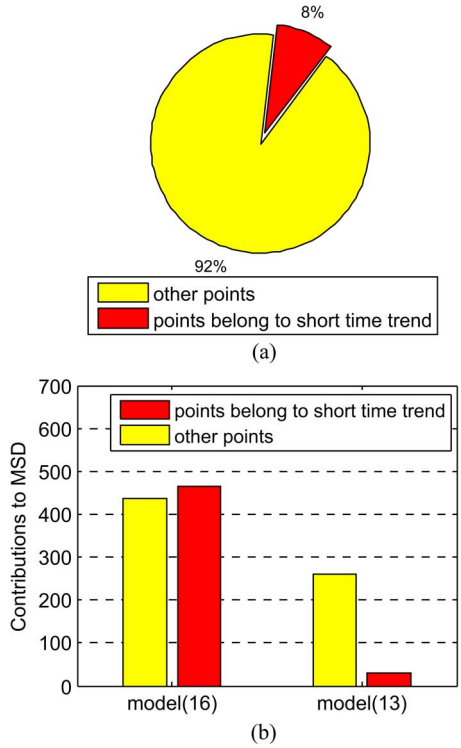


Fig. 10. (a) the ratios of the points that belong to short-term trends and the remaining points in the whole test dataset used in Fig. 9; (b) A comparison of prediction errors for short-term trends and the other points made by model (16) and model (13), respectively.

VI. Conclusion

In this paper, we revisit four important areas in traffic time series analysis: traffic prediction, data compression, abnormal data detection and missing data imputation. We show that for the studies on these subjects to yield meaningful results, appropriate modeling of traffic time series is needed. We also discover that each of aforementioned subjects mainly relies on one or two components of road traffic time series. Therefore, excluding unrelated components in forecast models has minimal negative effects on prediction performance.

Furthermore, we show that by using detrending methods, traffic time series can be decomposed into roughly independent components. More precisely, we find that:

1) The simple average trend usually reflects time-invariant feature of traffic time series; and the remaining components as a whole reflect time-variant fluctuations of traffic time series. Consequently, abnormal traffic scenarios can be detected by checking their trends, and feeding the residual time series rather than the original time series to prediction models will help magnify the fluctuation characteristics of traffic time series and thus lead to more accurate predictions.
2) The PCA trend provides a useful tool to simultaneously characterize the similarity and dissimilarity (variety) between observed daily traffic time series vectors. Moreover, the PCA trend is built upon the assumption that the dependences exist between data observed in different days/locations. Tests show that this assumption often

holds true for traffic time series, which makes PCA trend modeling particularly effective for traffic data compression and missing traffic data imputation applications.

3) The wavelet trend emphasizes the smooth variety tendency among consecutive data. It may be used to improve prediction performance.
4) The short-term trends within the residual time series reflect the temporary yet considerable dynamics of traffic time series. Using the information of multiple sensors will help prediction models track short-term trends and make more accurate predictions.

All these findings indicate that trend modeling is not only a technique to specify the temporal pattern, but is also related to the spatial relation of traffic time series. This conclusion justifies the necessity of developing an integrated framework for traffic time series analysis.

It must be pointed out that there are several other models (e.g., tensor decomposition model) to simultaneously model temporal trends and spatially joint occurrence dependences among traffic time series [51]–[54]. Due to page limit, we do not discuss these approaches in detail.

Lately, some multi-regime based traffic prediction [55]–[59] approaches have emerged. These approaches assume that traffic flow dynamics contain distinct patterns within a certain time interval and we should divide each daily traffic time series into different regimes based on these patterns. The jumps between regimes are often assumed to be abrupt [55]–[58] so that the obtained trend curve is usually piecewise-linear. Readers can find more discussions on the benefits and drawbacks of such multi-regime decomposition in our recent report [58]. There also exist models allowing smooth transition between regimes [59]. Constrained by length limit, we do not discuss them here.

Till now, it is yet unknown which model gives the best trend modeling results. We expect more attentions to be paid to this interesting yet important problem.

REFERENCES

[1] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerging Technol.*, vol. 43, pp. 3–9, Jun. 2014.

[2] C. Chen, J. Kwon, J. Rice, A. Skabardonis, and P. Varaiya, "Detecting errors and imputing missing data for single loop surveillance systems," *Transp. Res. Rec.*, no. 1855, pp. 160–167, 2002.

[3] X. Jin, Y. Zhang, L. Li, and J. Hu, "Robust PCA based abnormal traffic flow pattern isolation and loop detector fault detection," *Tsinghua Sci. Technol.*, vol. 13, no. 6, pp. 829–835, Dec. 2008.

[4] J.-M. Chiou, Y.-C. Zhang, W.-H. Chen, and C.-W. Chang, "A functional data approach to missing value imputation and outlier detection for traffic flow data," *Transportmetrica B, Transp. Dyn.*, vol. 2, no. 2, pp. 106–129, May 2014.

[5] S. Yang, K. Kalpakis, and A. Biem, "Detecting road traffic events by coupling multiple timeseries with a nonparametric Bayesian method," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1936–1946, Oct. 2014.

[6] J. Guo, W. Huang, and B. M. Williams, "Real time traffic flow outlier detection using short-term traffic conditional variance prediction," *Transp. Res. C, Emerging Technol.*, vol. 50, pp. 160–172, Jan. 2015.

[7] T. Tsekeris and A. Stathopoulos, "Measuring variability in urban traffic flow by use of principal component analysis," *J. Transp. Stat.*, vol. 9, no. 1, pp. 49–62, 2006.

[8] A. Stathopoulos and T. Tsekeris, "Methodology for processing archived ITS data for reliability analysis in urban networks," *Proc. Inst. Elect. Eng.—Intell. Transp. Syst.*, vol. 153, no. 1, pp. 105–112, Mar. 2006.

[9] L. Qu, J. Hu, and Y. Zhang, "A flow volumes data compression approach for traffic network based on principal component analysis," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2007, pp. 125–130.

[10] Z. Li, Y. Zhu, H. Zhu, and M. Li, "Compressive sensing approach to urban traffic sensing," in *Proc. 31st Int. Conf. Distrib. Comput. Syst.*, 2011, pp. 889–898.

[11] M. T. Asif, S. Kannan, J. Dauwels, and P. Jaillet, "Data compression techniques for urban traffic data," in *Proc. IEEE Symp. Ser. Comput. Intell.*, 2014, pp. 44–49.

[12] D. W. Xu, H. H. Dong, H. J. Li, L. M. Jia, and Y. J. Feng, "The estimation of road traffic states based on compressive sensing," *Transportmetrica B, Transp. Dyn.*, vol. 3, no. 2, pp. 131–152, May 2015.

[13] N. Mitrovic, M. T. Asif, U. Rasheed, J. Dauwels, and P. Jaillet, "CUR decomposition for compression and compressed sensing of large-scale traffic data," in *Proc. 16th Int. IEEE Annu. Conf. Intell. Transp. Syst.*, 2013, pp. 1475–1480.

[14] M. Zhong, S. Sharma, and Z. Liu, "Assessing robustness of imputation models based on data from different jurisdictions: Examples of Alberta and Saskatchewan, Canada," *Transp. Res. Rec.*, no. 1917, pp. 116–126, 2005.

[15] D. Ni, J. D. Leonard, II, A. Guin, and C. Feng, "Multiple imputation scheme for overcoming the missing values and variability issues in ITS data," *ASCE J. Transp. Eng.*, vol. 131, no. 12, pp. 931–938, Dec. 2005.

[16] L. Qu, Y. Zhang, J. Hu, L. Jia, and L. Li, "A BPCA based missing value imputing method for traffic flow volume data," in *Proc. IEEE Intell. Veh. Symp.*, 2008, pp. 985–990.

[17] L. Qu, L. Li, Y. Zhang, and J. Hu, "PPCA-Based missing data imputation for traffic flow volume: A systematical approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 3, pp. 512–522, Sep. 2009.

[18] W. Yin, P. M. Tuite, and H. Rakha, "Imputing erroneous data of single-station loop detectors for nonincident conditions: Comparison between temporal and spatial methods," *J. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 159–176, Jul. 2012.

[19] L. Li, Y. Li, and Z. Li, "Efficient missing data imputing for traffic flow by considering temporal and spatial dependence," *Transp. Res. C, Emerging Technol.*, vol. 34, pp. 108–120, Sep. 2013.

[20] Y. Li, Z. Li, and L. Li, "Missing traffic data: Comparison of imputation methods," *IET Intell. Transp. Syst.*, vol. 8, no. 1, pp. 51–57, Feb. 2014.

[21] M. T. Asif, N. Mitrovic, L. Garg, J. Dauwels, and P. Jaillet, "Low-dimensional models for missing data imputation in road networks," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 3527–3531.

[22] M. T. Asif *et al.*, "Spatial and temporal patterns in large-scale traffic speed prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 797–804, Apr. 2014.

[23] N. Mitrovic, M. T. Asif, J. Dauwels, and P. Jaillet, "Compressed prediction of large-scale urban traffic," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2014, pp. 5984–5988.

[24] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. C, Emerging Technol.*, vol. 10, no. 4, pp. 303–321, Aug. 2002.

[25] R. Chrobok, O. Kaumann, J. Wahle, and M. Schreckenberg, "Different methods of traffic forecast based on real data," *Eur. J. Oper. Res.*, vol. 155, no. 3, pp. 558–568, Jun. 2004.

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

[27] C. Chen, Y. Wang, L. Li, J. Hu, and Z. Zhang, "The retrieval of intra-day trend and its influence on traffic prediction," *Transp. Res. C, Emerging Technol.*, vol. 22, pp. 103–118, Jun. 2012.

[28] I. G. Guardiola, T. Leon, and F. Mallor, "A functional approach to monitor and recognize patterns of daily traffic profiles," *Transp. Res. B, Methodol.*, vol. 65, pp. 119–136, Jul. 2014.

[29] PeMS, California Performance Measurement System. [Online]. Available: http://pems.dot.ca.gov/

[30] G. W. Stewart, "On the early history of the singular value decomposition," *SIAM Rev.*, vol. 35, no. 4, pp. 551–566, Dec. 1993.

[31] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York, NY, USA: Springer-Verlag, 2002.

[32] P. Chaovalit, A. Gangopadhyay, G. Karabatis, and Z. Chen, "Discrete wavelet transform-based time series analysis and mining," *ACM Computing Surveys*, vol. 43, no. 2, p. 6, Jan. 2011.

[33] X. Jiang and H. Adeli, "Wavelet packet-autocorrelation function method for traffic flow pattern analysis," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 19, no. 5, pp. 324–337, Sep. 2004.

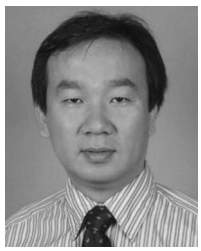
[34] X. Jiang and H. Adeli, "Dynamic wavelet neural network model for traffic flow forecasting," *ASCE J. Transp. Eng.*, vol. 131, no. 10, pp. 771–779, Oct. 2005.
[35] Y. Xie, Y. Zhang, and Z. Ye, "Short-term traffic volume forecasting using Kalman filter with discrete wavelet decomposition," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 22, no. 5, pp. 326–334, Jul. 2007.
[36] Y. Chen, Y. Zhang, J. Hu, and L. Li, "Mining for similarities in urban traffic flow using wavelets," in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2007, pp. 119–124.
[37] S. Dunne and B. Ghosh, "Weather adaptive traffic prediction using neurowavelet models," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 370–379, Mar. 2013.
[38] C. Chen, J. Hu, Y. Wand, Q. Meng, and Y. Zhang, "Short-time traffic flow prediction with ARIMA-GARCH model," in *Proc. IEEE Intell. Veh. Symp.*, 2011, pp. 607–612.
[39] Y. Zhang, Y. Zhang, and A. Haghani, "A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model," *Transp. Res. C, Emerging Technol.*, vol. 43, pp. 65–78, Jun. 2014.
[40] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transp. Res. C, Emerging Technol.*, vol. 43, Part 1, pp. 50–64, Jun. 2014.
[41] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys*, vol. 41, no. 3, p. 15, Jul. 2009.
[42] H. Teng and Y. Qi, "Application of wavelet technique to freeway incident detection," *Transp. Res. C, Emerging Technol.*, vol. 11, no. 3/4, pp. 289–308, Feb. 2003.
[43] Y.-S. Jeong, M. Castro-Neto, M.-K. Jeong, and L. D. Han, "A wavelet-based freeway incident detection algorithm with adapting threshold parameters," *Transp. Res. C, Emerging Technol.*, vol. 19, no. 1, pp. 1–19, Feb. 2011.
[44] W. J. Krzanowski, "Between-Groups comparison of principal components," *J. Amer. Stat. Assoc.*, vol. 74, no. 367, pp. 703–707, Sep. 1979.
[45] M. Hubert, P. Rousseeuw, and K. V. Branden, "ROBPCA: A new approach to robust principal components analysis," *Technometrics*, vol. 47, no. 1, pp. 64–79, Feb. 2005.
[46] Z. Zhao, Y. Zhang, J. Hu, and L. Li, "Comparison study of PCA and ICA based traffic flow compression," *J. Highway Transp. Res. Develop.*, vol. 4, no. 1, pp. 98–102, Jun. 2009.
[47] J. V. Stone, *Independent Component Analysis: A Tutorial Introduction*. Cambridge, MA, USA: MIT Press, 2004.
[48] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J. Roy. Stat. Soc. Ser. B, Stat. Methodol.*, vol. 21, no. 3, pp. 611–622, 1999.
[49] L. Li, X. Su, Y. Wang, Z. Li, and Y. Li, "Robust causality dependence mining in big data network and its application to traffic flow prediction," *Transp. Res. C, Emerging Technol.*, to be published. [Online]. Available: http://dx.doi.org/10.1016/j.trc.2015.03.003
[50] X. Zeng and Y. Zhang, "Development of recurrent neural network considering temporal-spatial input dynamics for freeway travel time modeling," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 28, no. 5, pp. 359–371, May 2013.
[51] H. Tan *et al.*, "A tensor-based method for missing traffic data completion," *Transp. Res. C, Emerging Technol.*, vol. 28, pp. 15–27, Mar. 2013.
[52] J.-M. Chiou, "Dynamic functional prediction and classification with application to traffic flow prediction," *Ann. Appl. Stat.*, vol. 6, no. 4, pp. 1588–1614, Dec. 2012.
[53] X.-Y. Huang *et al.*, "Multi-matrices factorization with application to missing sensor data imputation," *Sensors*, vol. 13, no. 11, pp. 15172–15186, Nov. 2013.
[54] B. Ran *et al.*, "Estimating missing traffic volume using low multilinear rank tensor completion," *J. Intell. Transp. Syst.*, to be published. [Online]. Available: http://dx.doi.org/10.1080/15472450.2015.1015721
[55] M. Cetin and G. Comert, "Short-term traffic flow prediction with regime switching models," *Transp. Res. Rec.*, no. 1965, pp. 23–31, 2006.
[56] Y. Kamarianakis, W. Shen, and L. Wynter, "Real-time road traffic forecasting using regime-switching space-time models and adaptive lasso," *Appl. Stochastic Models Bus. Ind.*, vol. 28, no. 4, pp. 297–315, Jul./Aug. 2012.
[57] C. Gurcan and B. Anton, "An online change-point-based model for traffic parameter prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1360–1369, Sep. 2013.
[58] Z. Li, Y. Li, and L. Li, "A comparison of detrending models and multi-regime models for traffic flow prediction," *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 4, pp. 34–44, Winter 2014.
[59] Y. Kamarianakis, H. O. Gao, and P. Prastacos, "Characterizing regimes in daily cycles of urban traffic using smooth-transition regressions," *Transp. Res. C, Emerging Technol.*, vol. 18, no. 5, pp. 821–840, Mar. 2010.

Li Li (S'05–M'06–SM'11) is currently an Associate Professor with the Department of Automation, Tsinghua University, Beijing, China, working in the fields of complex and networked systems, intelligent control and sensing, intelligent transportation systems, and intelligent vehicles. He is also with Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Nanjing, China. He is the author of more than 50 peer-reviewed journal papers and over 40 peer-reviewed conference papers as a first/corresponding author. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.

Xiaonan Su received the B.S.E.E. degree from Beijing Institute of Technology, Beijing, China, in 2012. He is currently working toward the Master's degree with the Department of Automation, Tsinghua University, Beijing.

Yi Zhang (S'01–M'04) received the B.S.E.E. and M.S. degrees from Tsinghua University, Beijing, China, in 1986 and 1988, respectively, and the Ph.D. degree in control science and engineering from the University of Strathclyde, Glasgow, U.K., in 1995. He is currently a Professor with Tsinghua University. He is also with Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Nanjing, China. His main research interests include advanced control theory and applications, advanced detection and measurement, and systems engineering. His current research interests include intelligent transportation systems, particularly intelligent vehicle–infrastructure cooperative systems, analysis of urban transportation systems, urban road network management, traffic data fusion and dissemination, and urban traffic control and management.

Yuetong Lin (M'05–SM'12) received the Ph.D. degree in systems and industrial engineering from the University of Arizona, Tucson, AZ, USA, in 2005. Since 2005, he has been with the Indiana State University, Terre Haute, IN, USA, where he is currently an Associate Professor of electronics and computer engineering technology.



Zhiheng Li (M'05) received the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2009. He is currently an Associate Professor with Department of Automation, Tsinghua University, Beijing, and with the Graduate School at Shenzhen, Tsinghua University, Shenzhen, China. His research interests include traffic operation, advanced traffic management system, urban traffic planning and intelligent transportation systems. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS.