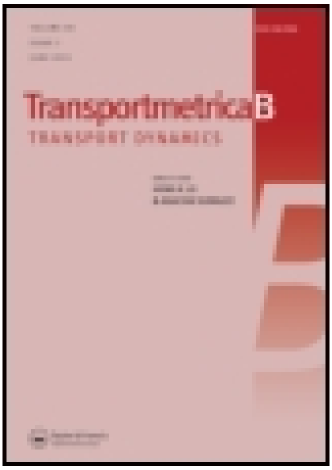

Transportmetrica B: Transport Dynamics

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ttrb20

A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD-ARIMA framework

Haizhong Wang[a], Lu Liu[a], Shangjia Dong[a], Zhen Qian[b] & Heng Wei[c]

[a] School of Civil & Construction Engineering, Oregon State University, Corvallis, OR 97331, USA

[b] H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

[c] School of Advanced Structures, University of Cincinnati, Cincinnati, OH 45221-0071, USA

Published online: 10 Jul 2015.

To cite this article: Haizhong Wang, Lu Liu, Shangjia Dong, Zhen Qian & Heng Wei (2015): A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD-ARIMA framework, Transportmetrica B: Transport Dynamics, DOI: 10.1080/21680566.2015.1060582

To link to this article: http://dx.doi.org/10.1080/21680566.2015.1060582

Transportmetrica B: Transport Dynamics, 2015
http://dx.doi.org/10.1080/21680566.2015.1060582

Taylor & Francis
Taylor & Francis Group

A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD–ARIMA framework

Haizhong Wang[a*], Lu Liu[a], Shangjia Dong[a], Zhen Qian[b] and Heng Wei[c]

[a]*School of Civil & Construction Engineering, Oregon State University, Corvallis, OR 97331, USA;* [b]*H. John Heinz III College, Carnegie Mellon University, Pittsburgh, PA, 15213, USA;* [c]*School of Advanced Structures, University of Cincinnati, Cincinnati, OH 45221-0071, USA*

(*Received 9 December 2014; accepted 7 June 2015*)

This paper presents a hybrid short-term traffic speed prediction framework through empirical mode decomposition (EMD) and autoregressive integrated moving average (ARIMA). The goals of this paper are to investigate (1) does the hybrid model provide better short-term traffic conditions (i.e. traffic speeds) than the traditional models? (2) how the performance of the hybrid model varies for varying scenarios such as mixed traffic flow and vehicle-type specific traffic prediction in a work zone, on-ramp, and off-ramp; and (3) why hybrid models provide better prediction than other single-staged models. Using empirical data from a work zone on interstate I91 in Springfield, MA and the on/off-ramp data from the Georgia State Route 400, the proposed hybrid EMD–ARIMA modelling framework is tested in the four distinct scenarios aforementioned. The prediction results of the hybrid EMD–ARIMA model are evaluated against the experimental data and also compared with the results from the traditional ARIMA, the Holt–Winters, the artificial neural network models, and a naive model. The evaluation results showed that the hybrid EMD–ARIMA model outperforms the traditional forecasting models in different scenarios.

Keywords: short-term traffic prediction; work zone; vehicle-type specific; empirical mode decomposition; ARIMA; EMD–ARIMA

1. Introduction

1.1. *Background*

The short-term traffic speed prediction is of paramount significance to forecast future traffic conditions within short time intervals such as 5 or 10 min. It can be used to indicate traffic status updates on a real-time basis. Reliable short-term traffic speed prediction holds a fundamental premise to enhance the efficacy of freeway operation and management via the Intelligent Transportation Systems (ITS) framework. Given availability of accurate traffic speed prediction, the Advanced Traveller Information System (ATIS) could be enhanced to provide the travelling public with more accurate travel time estimation on the basis of which travellers can arrange their schedule accordingly (Hamad et al. 2009). On the other hand, traffic operation decision-makers can also use the prediction results to deploy various traffic management

*Corresponding author. Email: haizhong.wang@oregonstate.edu

strategies through Advanced Traffic Management System to increase control and operational efficiency of the existing transportation network as a whole. While emerging advanced mobile information and communication technologies have implied a great potential to generate accurate and traceable traffic data, the challenge remains yet in modelling a reliable short-term speed prediction function under the ITS circumstance.

In general, there are five timescale requirements for prediction models (Giebel and Kariniotakis 2009) and the prediction scale is categorised into the following spectrums: ultra short-term range (in seconds), very short-term range (1–10 minutes to an hour), short-term range (0–6 or 8 h), medium-term range (0 hour to 7 days ahead), and long-term range (weeks). Statistical methods (i.e. the autoregressive integrated moving average (ARIMA) model) and artificial intelligence (AI) based methods (i.e. artificial neural network (ANN)) have been conventionally used by researchers and practitioners to deal with traffic prediction modelling problems. However, the linear assumption of the ARIMA model and the 'black box' nature of the ANNs method suggests an obvious limitation as opposed to high nonlinearities and irregularities of the traffic speed series in reality. Such an assumption incompatibility issue in applying the statistical methods or AI-based ANN methods could not be easily recognised by the users, because their modelling structures are hidden in a 'black box' effect as the models are usually wrapped within a commercial computing software package. To overcome the aforementioned limitation, a hybrid empirical mode decomposition (EMD) and ARIMA model, or the hybrid EMD–ARIMA approach, has been developed to predict short-term freeway traffic speed. With the hybrid EMD–ARIMA approach, the EMD model is used to decompose the original series, with its capability to deal with nonlinear and non-stationary series. Compared with AI-based methods, this method favours a more transparent and explicit structure.

The empirical data used in the case study were collected through the Georgia ITS program of the Georgia Department of Transportation on the State Route 400 and a work zone from Interstate 91 in Springfield, MA. In the case of missing data in empirical study, researches have been done to address the problem. Chiou et al. (2014) proposed to impute missing values by using the conditional expectation approach to functional principal component analysis (FPCA). They proved that FPCA performs better than the probabilistic principal component analysis (PCA) and the Bayesian PCA. Due to the large numbers of injuries and fatalities involved with work zone crashes, researchers and practitioners have been trying to build design guidelines for work zones (Ullman et al. 2004; Wunderlich 2005). At the federal level, the design guidance is provided by MUTCD, state departments of transportation are also developing those guidelines (Administration 2003). Also, researchers have been studying the capacities and speed profiles of work zones. However, that research mainly focused on the influence to freeway capacities with the presence of work zone and the relationship between speed and the work zone geometric characteristics. The study on work zone traffic speed forecast, which can give us a better understanding of the speed profile and safety effects, is limited. To the best of our knowledge, Taylor et al. (2007) built a neural network model to predict the 15th, mean, and 85th speed. They concluded that ANN is able to predict the vehicle operating speeds in work zones with reasonable accuracy.

1.2. *Objective of paper*

This paper implements a hybrid model through EMD and ARIMA to predict short-term freeway traffic speed. The proposed hybrid EMD–ARIMA modelling framework is tested on different vehicle types speed prediction in three distinct scenarios: work zone environment, on-ramp, and off-ramp. The results showed that the hybrid EMD–ARIMA model outperformed the traditional time series forecasting models (i.e. Holt–Winters or ARIMA itself), ANN-based models (i.e. ANN) and a naive model. The major contributions of this research are threefold: (1) the hybrid EMD–ARIMA modelling framework is proposed and tested in four distinct scenarios: mixed traffic in a work zone, vehicle-type specific prediction in a work zone, on-ramp, and off-ramp; (2) the prediction results of the proposed model is evaluated qualitatively and numerically which showed that the hybrid model outperformed the traditional ARIMA, the Holt–Winters, the ANN model, and the simple model; (3) the experimental results showed that the performance of the hybrid modelling framework is consistent with varying traffic conditions in different scenarios.

1.3. *Paper organisation*

The remainder of this paper is organised as follows: a literature review on different models for short-term forecasting is first presented in Section 2. Section 3 presents the framework of the hybrid EMD–ARIMA model and the detailed modelling steps of the EMD method and ARIMA model. Section 4 provides descriptions of the empirical data and the study site in this research. Section 5 displays three traffic speed forecasting cases (freeway basic segment, on-ramp and on-ramp segment) for single and multiple-step forecasting horizon (i.e. 5, 10, 15 and 20 min). Finally, the conclusions are stated in Section 6 and future research work is also discussed.

2. Literature review

Over the past decades, researchers and practitioners have developed numerous predication methods which can be categorised as (1) statistical models; (2) AI-based models; and (3) hybrid forecasting models (Vlahogianni, Golias, and Karlaftis 2004). Statistical models generally include a spectrum of time series models such as ARIMA models to predict traffic flow (i.e. volume and speed) (Van Der Voort, Dougherty, and Watson 1996; Lee and Fambro 1999; Koutroumanidis, Ioannou, and Arabatzis 2009), parametric and non-parametric models (Smith, Williams, and Oswald 2002), multivariate state space models for traffic flow prediction (Stathopoulos and Karlaftis 2003), the nonlinear dynamic approach to study temporal evolution of traffic (Vlahogianni, Karlaftis, and Golias 2008), and Bayesian-based prediction mechanisms. While the AI-based prediction methods include: ANN (Smith and Demetsky 1994; Chang and Su 1995; Innamma 2005; Taylor et al. 2007), fuzzy logic methods for short-term traffic flow forecasting (Ye and Zhang 2008), neuron-genetic algorithm for traffic flow prediction (Abdulhai, Porwal, and Recker 2002), Kalman filtering methods to predict traffic volume and estimate traffic states (Okutani and Stephanedes 1984; Xie, Zhang, and Ye 2007; Papageorgiou and Wang 2007), and support vector machine (SVM) for online short-term traffic flow prediction (Castro-Neto et al. 2009). Hybrid forecasting models typically combine two or more models to form synergetic prediction models. Some examples of hybrid models are EMD–ANN to predict wind speed (Liu et al. 2012), EMD–SVM for air passenger flow prediction (Bao, Xiong, and Hu 2012), ARIMA–SVM to predict stock and carbon prices (Pai and Lin 2005; Zhu and Wei 2013), ARIMA–EGARCH (Kumar and Thenmozhi 2012), EEMD–ANN, etc. Literature about these methods to various applications abounds. Researchers and practioners have been trying to better understand the pros and cons of each method, for example, Watson, Kirby, and Dougherty (1997) presented an extensive review about which method should be used for short-term traffic forecasting between neural network and statistical methods. Mori et al. (2014) conducted a thorough review on the modelling of travel time. It focuses on travel time estimation and travel time prediction. In general, the community has relatively better understandings about the individual models than hybrid methods. The authors summarised the existing relevant short-term traffic forecasting studies in the three categories aforementioned in Figure 1. However, this review is not intended to be complete in any sense.

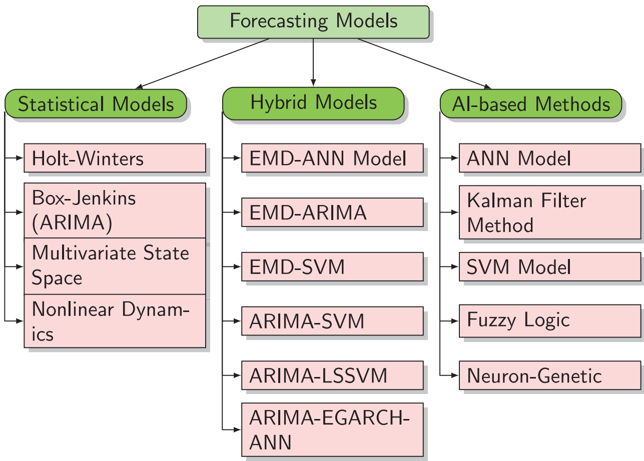

Figure 1. Summary of available forecasting models.

2.1. *Statistical models*

The time-series-based statistical forecasting methods cover a wide spectrum of models starting from moving average and simple regression techniques (including parametric and non-parametric) to relatively sophisticated methods such as the Holt–Winters forecasting procedure (Chatfield 1978) and Box–Jenkins methodology (Box and Jenkins 1976). Additionally, Moorthy and Ratcliffe (1988) explored the application of time-series analysis to produce short-term forecasts and investigated the performance of the forecasting models. Lee and Fambro (1999) investigated the use of the ARIMA model for short-term traffic volume forecasting, which showed that the use of a subset of the ARIMA model would increase the accuracy of the short-term forecasting performance within time-series models. Van Der Voort, Dougherty, and Watson (1996) introduced a hybrid short-term forecasting method known as the KARIMA method and used the Kohonen self-organising map as an initial classifier which showed improvement over the single ARIMA model. Williams, Durvasula, and Donald (1998) applied a seasonal ARIMA and exponential smoothing model to predict urban freeway traffic flow. Smith, Williams, and Oswald (2002) examined some non-parametric regression coupled with heuristic forecast generation methods against seasonal ARIMA and found that their performance cannot equal that of seasonal ARIMA. Stathopoulos and Karlaftis (2003) collected 3 min data from urban arterials streets near downtown Athens and used the data to estimate multivariate time-series models. They found that, ARIMA models provided the best fit that turned out to be quite useful especially for points of entry into the study area. Kamarianakis and Prastacos (2003) presented and discussed two different univariate (historical average and ARIMA), two different multivariate (i.e. vector autoregressive moving average (VARMA) and the space–time ARIMA (STARIMA)), and compared the performance of the four models. The results showed that the performance of ARIMA, VARMA and STARIMA models was comparable. Vlahogianni, Karlaftis, and Golias (2006) did some research on assessing the statistical properties of short-term univariate traffic volume series regarding overall nonlinearity and non-stationarity which was revealed by the surrogate data test using the time reversibility as a criterion. This work is of critical importance since the statistical forecast models always need some assumptions and tests about the statistical properties of the original time series. Further, Karlaftis and Vlahogianni (2009) studied the lack of ability to capture long memory properties in ARIMA and generalised autoregressive conditional heteroskedasticity models and employed a fractionally integrated dual memory model to compare with the classical time-series method.

2.2. *AI-based methods*

In general, the prediction methods aforementioned can provide reliable prediction results under specific circumstances. However, the traffic speed series is highly nonlinear and irregular which is contrary to the linear assumption of tradition prediction methods (Hamad et al. 2009; Bao, Xiong, and Hu 2012; Wei and Chen 2012). Due to the inherent limitation, many AI models, for example, artificial neural networks, SVM, or Kalman filter, are employed in the short-term forecasting research.

There has been a lot of discussions in the literature upon the use of ANN in short-term predictions. Smith and Demetsky (1994) and Dougherty and Kirby (1998) were among the early group of researchers who successfully applied neural network to short-term traffic flow forecasting such as traffic congestion prediction. Later, Jiang and Adeli (2005) and Innamma (2005) used neural network to predict the short-term traffic flow performance and demonstrated the capability to project future traffic conditions. Vlahogianni, Karlaftis, and Golias (2005) used a genetic approach to optimise the neural network and found that a genetically optimised network has much better ability to capture high flow values to overcome a major shortcoming of other tested approaches such as the ARIMA models. However, the neural network-based prediction mechanism has been criticised for its inherent 'black box and hidden structure' nature and intensive demand of training data although the prediction accuracy seems appealling for practical implementation purposes. Watson, Kirby, and Dougherty (1997) questioned the neural network approach and raised the questions regarding whether we should use statistical methods or neural networks for short-term traffic flow forecasting. In this study, the authors found that the 'black box' issue of neural network with their hidden structure needs to be considered seriously in the application of traffic-related predictions.

For other AI models, Castro-Neto et al. (2009) conducted online short-term traffic flow prediction through an SVM-based approach for normal and abnormal traffic conditions. As traffic speed forecasting is essential to the travel time reliability, Chen et al. (2013) addressed the problem of finding the least expected time path in stochastic time-dependent road networks. Okutani and Stephanedes (1984) proposed two models employing Kalman filtering theory for predicting short-term traffic volume and found that average prediction error is found to be less than 9% and the maximum error less than 30% compared with benchmarking models. Likewise, Xie, Zhang, and Ye (2007) suggested the approach of the Kalman filter and discrete wavelet decomposition. Later, Ye and Zhang (2008) performed short-term traffic flow forecasting through a fuzzy logic system approach. Finally, Papageorgiou and Wang (2007) proposed an extended Kalman filter-based approach to estimate real-time freeway traffic states.

2.3. *Hybrid forecasting models*

The forecasting results based on the AI-based models showed some advantages over the traditional methods; however, these AI methods also have their shortcomings. For example, the dimensionality of traffic data creates the challenge in the ANN approach, while genetic algorithm (GA) and SVM are quite sensitive in selecting parameters and the local minima. Overfitting has been a possible issue in statistical or AI-based models. In order to resolve these flaws, the hybrid methods are developed to address the short-term traffic forecasting and have been proved to be effective.

In recent forecasting literature, researchers applied the fusion of EMD, ANN, GA and SVM to make short-term predictions. Liu et al. (2012) and Wei and Chen (2012) used a synergitic EMD and ANN to predict wind speed and metro passenger flow. Bao, Xiong, and Hu (2012) proposed the hybrid EEMD–SVM approach to forecast the air traffic passenger flow using six airlines' empirical data. The results showed that the proposed hybrid modelling methodology outperforms the pure SVMs, Holt–Winters, and ARIMA. Pan et al. (2013) extended stochastic cell transmission model (SCTM) framework to incorporate the spatial–temporal correlation of traffic flow and to support short-term traffic state prediction. A multivariate normal distribution-based best linear predictor is used to forecast boundary variables and/or supply functions, the predicted results then serves as inputs to the SCTM to perform short-term traffic state prediction. Abdulhai, Porwal, and Recker (2002) trained the neural network with GAs to forecast the short-term freeway traffic flow. Hamad et al. (2009) investigated the near-term travel speed forecasting using a Hilbert–Huang transform (HHT) which is actually a hybrid model of EMD and the back-propagation neural network. Ngo, Apon, and Hoffman (2012) utilised several forecasting techniques (VAR, vector autoregressive and ARIMA) in combination with EMD to investigate the trade-offs of EMD's decomposition (sifting) step for forecasting the arrival workload of an enterprise cluster. The results showed that EMD helps to improve forecasting results. Okolobah and Ismail (2013) proposed a method based on EMD and ARIMA to predict the peak load and the result showed that the proposed model presented a better forecasting accuracy than the traditional ARIMA method. The primary motivations of these hybrid methods are to fix the individual flaws and generate a synergetic way in short-term prediction of wind speed, traffic flow, passenger flow, etc.

3. The hybrid EMD–ARIMA modelling framework

Inspired by the hybrid methods, this paper proposes a hybrid method of EMD–ARIMA for short-term traffic speed forecasting. In this methodology, we use the EMD technique to decompose the original traffic speed data into independent components. The main purpose of decomposition is twofold: (1) to simplify the forecasting and (2) to differentiate the modes with different characters (i.e. intrinsic mode functions (IMFs)) contributing to the prediction accuracy. The combined procedure is to develop a consensus prediction on immediate historical data. Utilising the EMD techniques, the original traffic speed series with nonlinear and non-stationary features were decomposed into finite and small number of IMFs. After these IMFs are obtained, each IMF is modelled by an ARIMA model and can be accurately forecasted. Finally, all the predicted results are aggregated to generate a combined forecasting result (Figure 2). Refer to Figure 3 for a detailed procedure of the EMD process.

The hybrid modelling framework through EMD–ARIMA includes two major steps: (1) EMD and (2) ARIMA. Within the hybrid EMD–ARIMA framework, the EMD model is used to decompose the original series, particularly in dealing with nonlinear and non-stationary series. Compared with AI-based methods, this method favours a more transparent and explicit structure. Traditionally, researchers have been using time–frequency analysis algorithms such as short-time Fourier transform or wavelet transform (WT) to decompose the original signal into a series of sub-series. The major difference between wavelet transform and EMD is WT uses wavelet basis (such as Harr, Morlet, Daubechies, etc.) to decompose the original signal while EMD decomposes signal through an iterative sifting process to extract the IMFs. EMD is particularly suitable for nonlinear and non-stationary signal while WT is more appropriate for linear signals (Huang et al. 1998). The essence of the EMD is to decompose the original signal into the sum of a finite number of IMFs and the mean trend (residual). The decomposition process of EMD is based on the local characteristic time scale to obtain the IMFs. This hybrid modelling framework has been applied to predict arrival data of an enterprise computing system (Ngo, Apon, and Hoffman 2012) and peak load demand forecasting (Okolobah and Ismail 2013) with reasonable success.

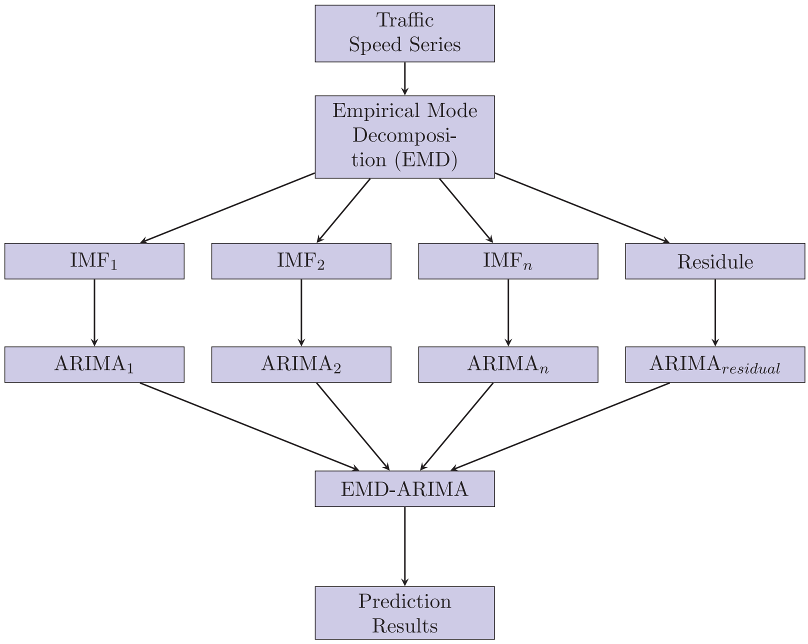

Figure 2. The hybrid EMD–ARIMA modelling framework.

3.1. *Stage 1: empirical mode decomposition*

Following the pioneering work of Huang et al. (1998) on the HHT, the EMD has been widely used to decompose a signal into the IMF and use the Hilbert Spectrum Analysis to analyse non-stationary and nonlinear time-series data. Empirical traffic speed series over time is non-stationary and nonlinear time series as can be seen from Figure 4. Due to the highly nonlinear and non-stationary nature of traffic speed series, by decomposing the time-series IMF, more accurate predictions would be obtained (Liu et al. 2012; Wei and Chen 2012).

Given an original traffic volume or speed series $x(t)$, it can be written into the following form after applying the EMD:

$$x(t) = r(t) + \sum_{i=1}^{n} d^i(t) \tag{1}$$

in which $r(t)$ is the residual after n IMFs are derived and $d^i(t),\ i = 1, 2, \ldots, n$ is the IMF for different decompositions. In the EMD, an IMF is defined as a function satisfying the following requirements:

(a) In the whole data set, the number of extrema and the number of zero-crossings must either be equal or differ at most by one.
(b) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

An IMF can have variable amplitude and frequency along the time axis. An iterative procedure called the sifting process is adopted to extract the separate IMF components. The detailed procedures of the sifting process can be summarised as follows:

(I) Identify all the local extrema (i.e. local maxima and minima) in the time series $x(t)$
(II) Interpolate all the local maxima by a cubic spline line to form the upper envelope $e_{\max}(t)$.

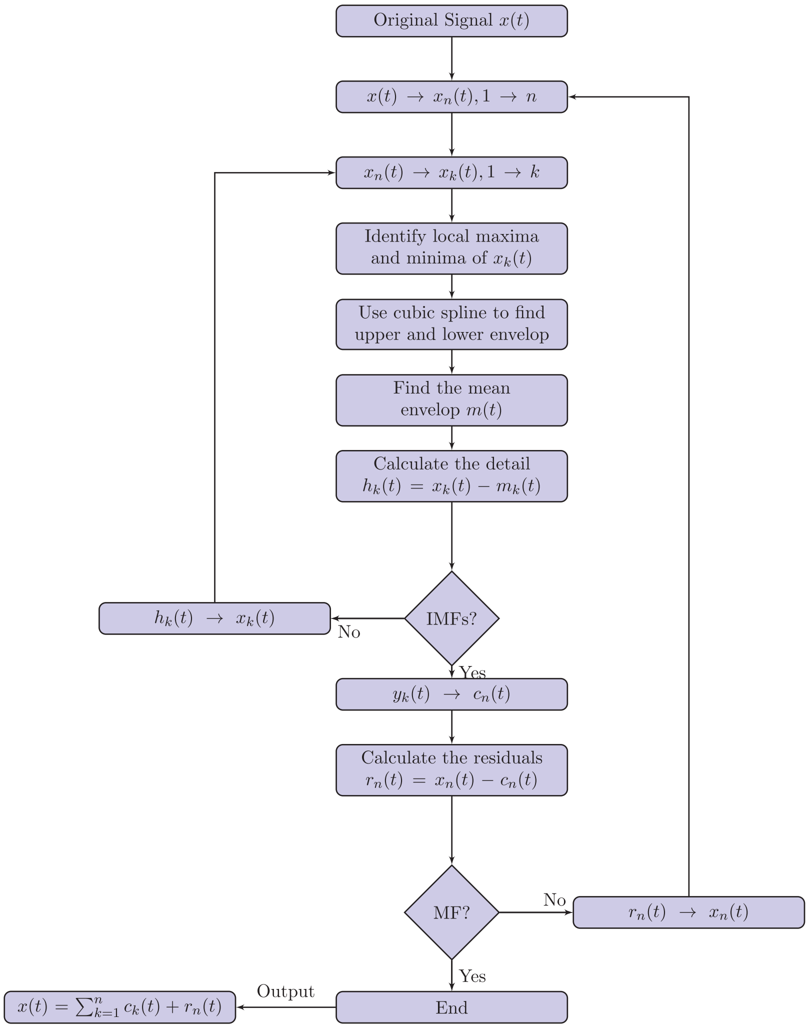

Figure 3. The flow chart of the EMD algorithm.

(III) Repeat the procedure for the local minima to produce the lower envelope $e_{\min}(t)$.

(IV) Compute the mean envelope $m(t)$ from the upper (high) and lower envelop as follows:

$$m(t) = \frac{[e_{\max}(t) + e_{\min}(t)]}{2}. \tag{2}$$

(V) Extract the mean envelope from the original signal to obtain a new component as follows:

$$h(t) = x(t) - m(t). \tag{3}$$

(VI) Check whether $h(t)$ is an IMF: (1) If $h(t)$ is an IMF then set $d(t) = h(t)$ and replace $x(t)$ with the residual $r(t) = x(t) - d(t)$; (2) if $h(t)$ is not an IMF, replace $x(t)$ with $z(t)$ then

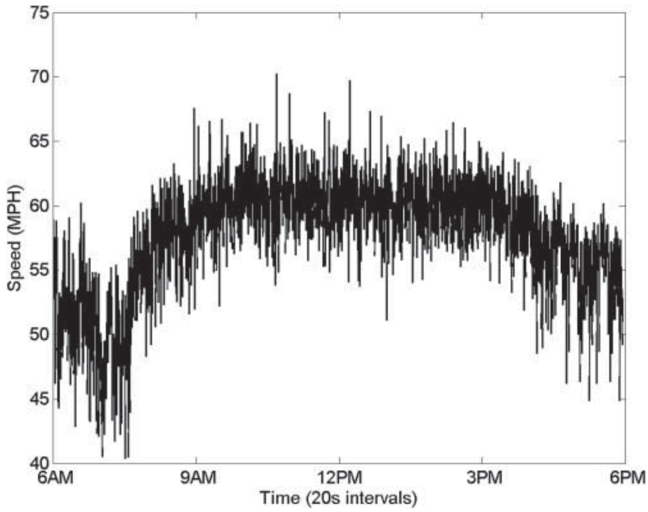

Figure 4. The traffic speed series over an one-day observation from the freeway basic segment.

repeat the steps from II to V until the following stopping criterion is met in the iterative process:

$$\mathrm{SD} = \sum_{t=1}^{m} \frac{[h_{i-1}(t) - h_i(t)]^2}{[h_{i-1}(t)]^2} \leq \delta \quad (i = 1, 2, \ldots, m). \tag{4}$$

The criterion for the sifting process to stop: SD (standard deviation), computed from the two consecutive sifting results is defined as above. A typical value for SD can be set between 0.2 and 0.3.

3.2. *Stage 2: ARIMA model building*

The ARIMA model has been widely used in time-series predictions since it was initially proposed by Box and Jenkins (1976). For example, it has been applied to predict traffic flow by Lee and Fambro (1999) and Williams, Durvasula, and Donald (1998). The results both showed using a subset of the ARIMA model would increase the accuracy of the short-term forecasting task within time-series models. In addition, a hybrid short-term forecasting method called KARIMA is developed by Van Der Voort, Dougherty, and Watson (1996) to show the improved performance over single ARIMA models. All of these show that the ARIMA model proved to be an effective method in short-term predictions.

To build an ARIMA model for each IMF, we start with identifying the correlation (sample autocorrelation and sample partial autocorrelation functions (PACFs)) using time-series analysis. The autocorrelation function (ACF) describes the correlation between values of a time series at different times as a function of the two times or the time lag. Let X be the time series, and i be some point in time after the start of the time series. Then X_i is the value at time i, suppose the time series has mean μ_i and variance σ_i^2, then the definition of the autocorrelation between time point s and t is:

$$R(s,t) = \frac{E[(X_t - \mu_t)(X_s - \mu_s)]}{\sigma_t \sigma_s}. \tag{5}$$

PACF describes the correlation between series values that are k intervals apart. Accounting for the values of the intervals between them, it is an important function in time-series analysis to identify the extent of the lag in an autoregressive model. Given a time series z_t, the partial autocorrelation of lag k is the autocorrelation between z_t and z_{t+k} with the linear dependence of z_{t+1} to z_{t+k-1} removed, we can define the partial correlation function as

$$\alpha(1) = \mathrm{Cor}(z_t, z_{t+1}), \tag{6}$$

$$\alpha(k) = \mathrm{Cor}(z_{t+k} - P_{t,k}(z_{t+k}), z_t - P_{t,k}(z_t)) \quad \text{for } k \geq 2, \tag{7}$$

where $P_{t,k}(x)$ is the projection of x onto the space spanned by $z_{t+1}, \ldots, z_{t+k-1}$. After the identification of ACF and PACF, the model order (p, d, q) can be determined, then by applying the least-square estimation, the parameters of the ARIMA model can be calculated.

4. Experiments

In this section, the authors show the empirical data sets used in this research. In addition, the model-building procedures are shown other than the proposed hybrid prediction framework.

4.1. *Data description*

The empirical data used in this research were collected from a work zone on interstate I91 in Springfield, MA and the on/off-ramps on the Georgia State Route 400 which is the South/North freeway entering/exiting the city of Atlanta. The data were initially collected through video cameras and the data were extracted from the videos through virtual loop detectors. The empirical data were collected over an approximately 20 mile long corridor including 100 detectors in which 78 detectors were installed on the basic freeway segments and 22 deployed on the on/off-ramps. The geographic site and the study area from I91 and from GA400 are given in Figure 5.

The traffic speed data used in this experiment are a 20 s data by aggregation including three distinct scenarios: the basic freeway segment, the on-ramp, and the off-ramp. The study time periods include morning/evening peak hours (7:00 am to 9:00 am and 4:00 pm to 6:00 pm) and non-peak hours over the day for five weekdays (02/10/2003 to 02/14/2003). The data from each day were divided into two subsets of data; one subset of data was used to build the hybrid EMD–ARIMA forecasting model, and the other subset was used as the empirical ground truth data to compare against the forecasting results. Figure 4 shows one day's empirical traffic speed observation from the basic freeway segment on GA400. The maximum speed of the one sampled day's data is 65.03 mph, the minimum speed is 29.91 mph, the average speed is 55.78 mph, and the SD is 4.93.

4.2. *Extraction of IMF components*

Following the EMD steps outlined in Section 3 Stage 1, a total of 10 imfs and 1 residue were obtained from the original speed data series, as is shown in Figure 6.

Figure 6 shows that the original speed series are decomposed into a series of relatively stationary speed data sets which will be modelled later by ARIMA. The short-period (or high frequency) components are extracted in the first few components, such as IMF 1 and IMF 2, and long-period (or low frequency) components are given in the last few components, such as IMF 9 and IMF 10. The highly stochastic variations in the speed series are represented by the first few components, while the cyclical components in the speed series are shown by the last few components. The last component (represented as Residue) is the residue of the sifting process which shows the trend of the speed series.

The original time series can be reconstructed by utilising the sifting process backward. By doing this, we can also examine the intrinsic meaning of these IMF components. The

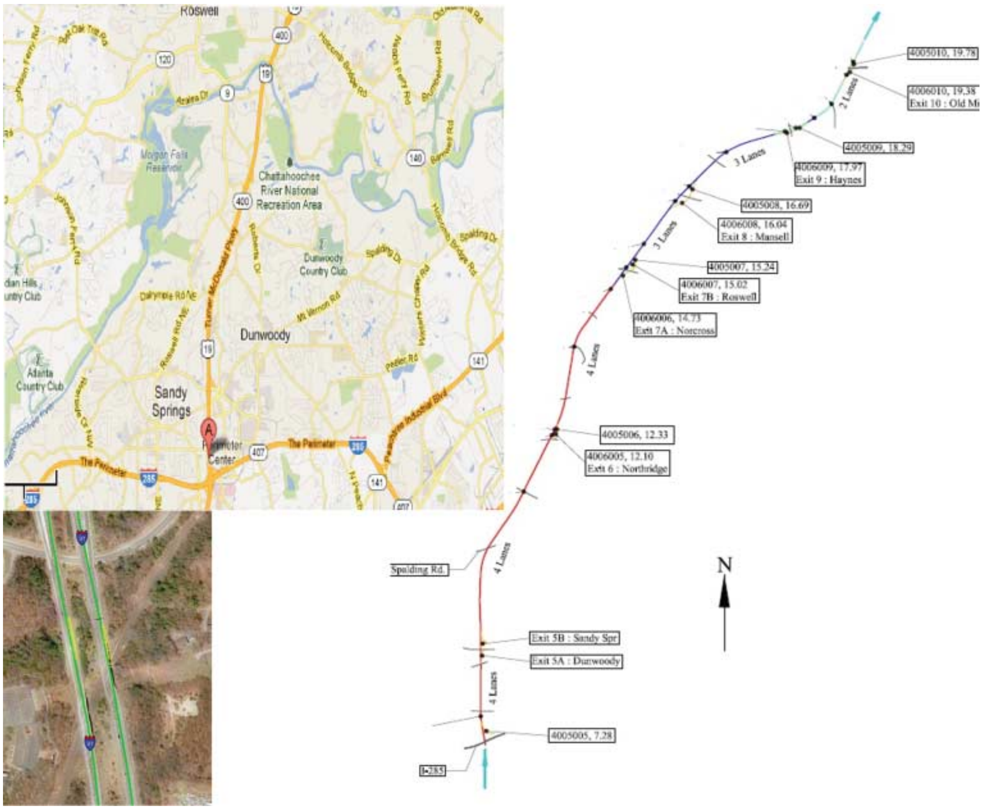

Figure 5. The selected study site from I91 in Springfield, MA and the on/off-ramps from GA400 Northbound at Atlanta, Georgia (Source: Google map and hand sketches).

short-period components are illustrated in the first few IMFs and long-period components are extracted in the last few components. The last component is the residue of the sifting process and it represents the trend of the original time series. The reconstruction process is illustrated in Figure 7.

The original traffic speed signal can be retrieved by the combination of each IMF component and residue. Interested readers are directed to Hamad et al. (2009) to learn how to retrieve the original signal from the decomposed IMFs.

4.3. *Building procedures of other models for comparison*

4.3.1. *The simple model*

A simple historical average model is added to provide a naive predictor in the different scenarios in the results section. In the simple model, the forecast is equal to the last observed value.

4.3.2. *ANN model*

For the ANN model, the authors used a feed forward neural network with a Levenberg–Marquardt training function. We have tried four training functions: Levenberg–Marquardt, Bayesian Regularisation, BFGS Quasi-Newton, and Resilient Backpropagation. Among them

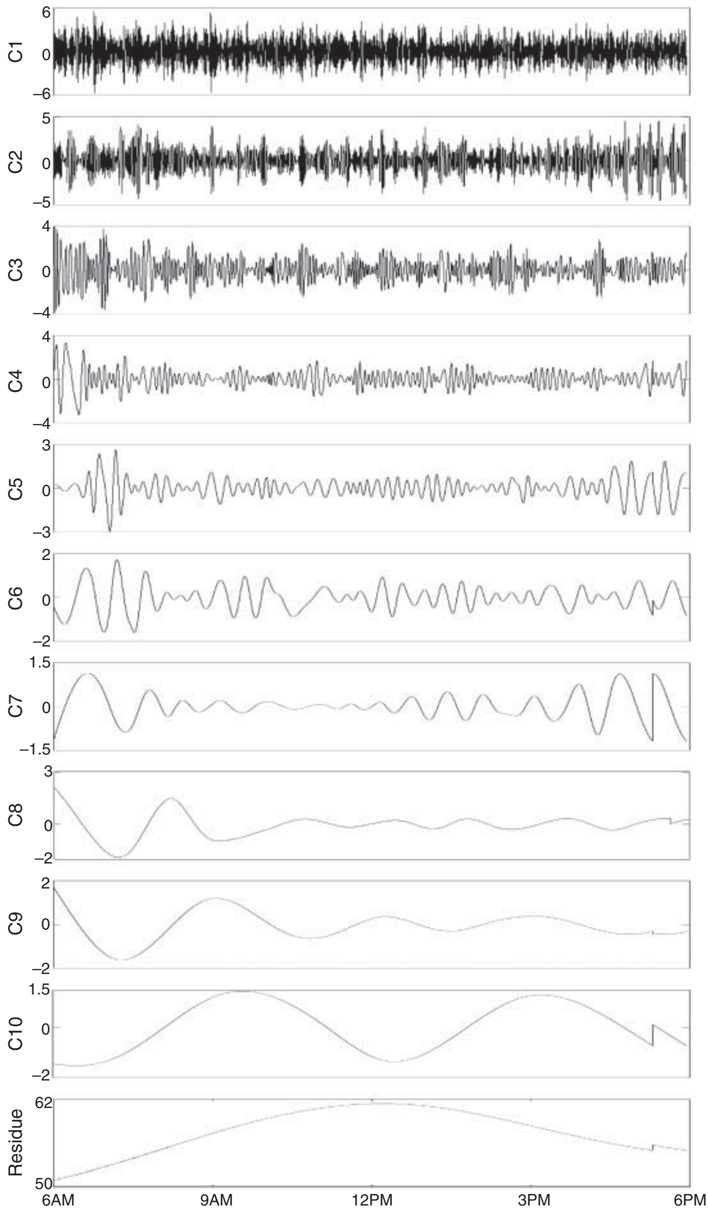

Figure 6. The IMFs and residue extracted through EMD.

the Levenberg–Marquardt performed the best. There are two parameters that need to be determined: number of past values(number of delays) and number of hidden layer nodes. The number of hidden layer nodes should be determined with caution because it can affect either the learning or forecast capability. Hidden layer nodes have to capture the dynamics of the speed series and the associated patterns. The number of time-lags of hidden layer nodes is based on the rolling step and the rolling horizon. The rolling step represents the resolution of the speed series used for forecast. The rolling horizon is equal to the number of time-lags of hidden layer

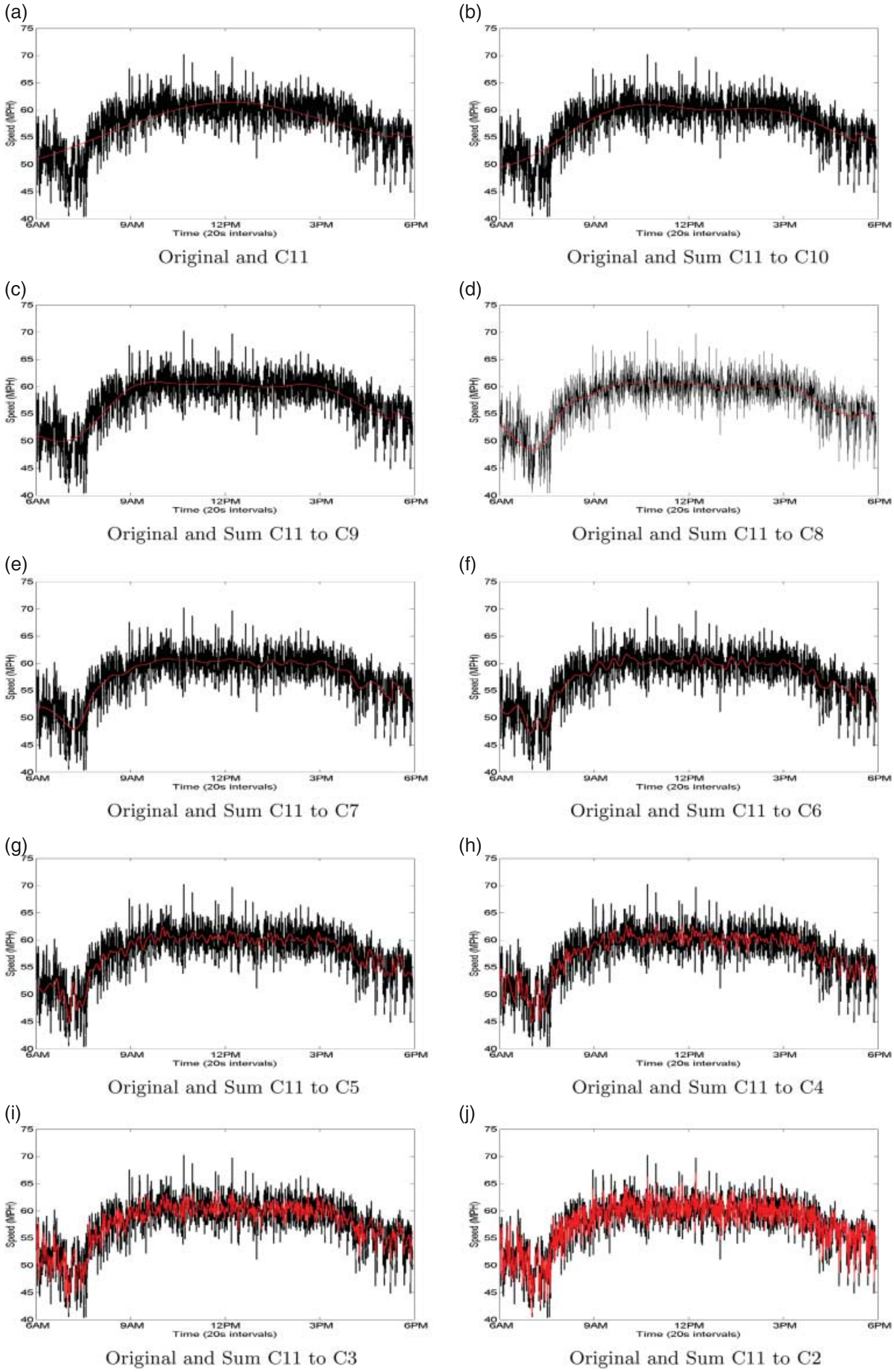

Figure 7. Reconstruction process of the original time series from IMFs.

nodes. If the rolling horizon is too large, the forecasting model responds slowly to the fluctuation of the speed series. On the contrary, if the rolling horizon is too small, the forecasting model overreacts to the fluctuation of the speed series. In the output layer, output variables include the forecasting step and the forecasting horizon. The forecasting step represents the time interval in which the forecast is performed. The forecasting horizon indicates the extent of time ahead.

In our experiment, we have tried a combination of (1) number of past values and (2) number of hidden layer nodes such as (10, 10), (10, 15), (10, 20), (15, 10), (15, 15), (15, 20), (20, 10), (20, 15), (20, 20). Finally, it turns out that the combination (15, 15) performed the best.

4.3.3. *The Holt–Winters model*

The forecast through the Holt–Winters exponential smoothing method is done through estimating the level, slope and seasonal components at the current time point. Smoothing is controlled by three parameters: α, β, and γ. The parameters α, β and γ all have values between 0 and 1. Values that are close to 0 mean that relatively little weight is placed on the most recent observations when making forecasts of future values. In the experiment of a one day traffic speed series, the smoothing parameters are $\alpha = 0.1487768$, $\beta = 0.01610834$, and $\gamma =$ FALSE (indicating that it is a non-seasonal series).

For the Holt–Winters exponential smoothing method, what we need to determine is the estimates of the level, slope and seasonal component at each time point. The forecast function of the Holt–Winters exponential smoothing method is as follows:

$$\hat{y}_{n+l|n} = (m_n + l \times b_n) \times c_{n-s+l} \quad l = 1, 2, \ldots, \tag{8}$$

where m_n is the component of level, b_n is the component of slope, and c_{n-s+l} is the relevant seasonal component with s representing the seasonal period.

The updating formula of the three components will require a smoothing constant each. If α_0 is used as the parameter for the level component and α_1 for the slope component. A third constant α_2 is added as the smoothing constant for the seasonal factor, the updating equations will be:

$$m_t = \alpha_0 \times \frac{y_t}{c_{t-s}} + (1 - \alpha_0) \times (m_{t-1} + b_{t-1}), \tag{9}$$

$$b_t = \alpha_1 \times (m_t - m_{t-1}) + (1 - \alpha_1) \times b_{t-1}, \tag{10}$$

$$c_t = \alpha_2 \times \frac{y_t}{m_t} + (1 - \alpha_2) \times c_{t-s}, \tag{11}$$

α_0, α_1 and α_2 lie between zero and one. With the updating formula, the seasonal factor is added as opposed to multiplied into the one step ahead forecast function:

$$\hat{y}_{n+l|n} = m_n + b_n + c_{n-s+l} \tag{12}$$

the updated level and seasonal equations involve differences as opposed to ratios:

$$m_t = \alpha_0 \times (y_t - c_{t-s}) + (1 - \alpha_0) \times (m_{t-1} + b_{t-1}), \tag{13}$$

$$c_t = \alpha_2 * (y_t - m_t) + (1 - \alpha_2) * c_{t-s} \tag{14}$$

and the slope component b_t will remain unchanged.

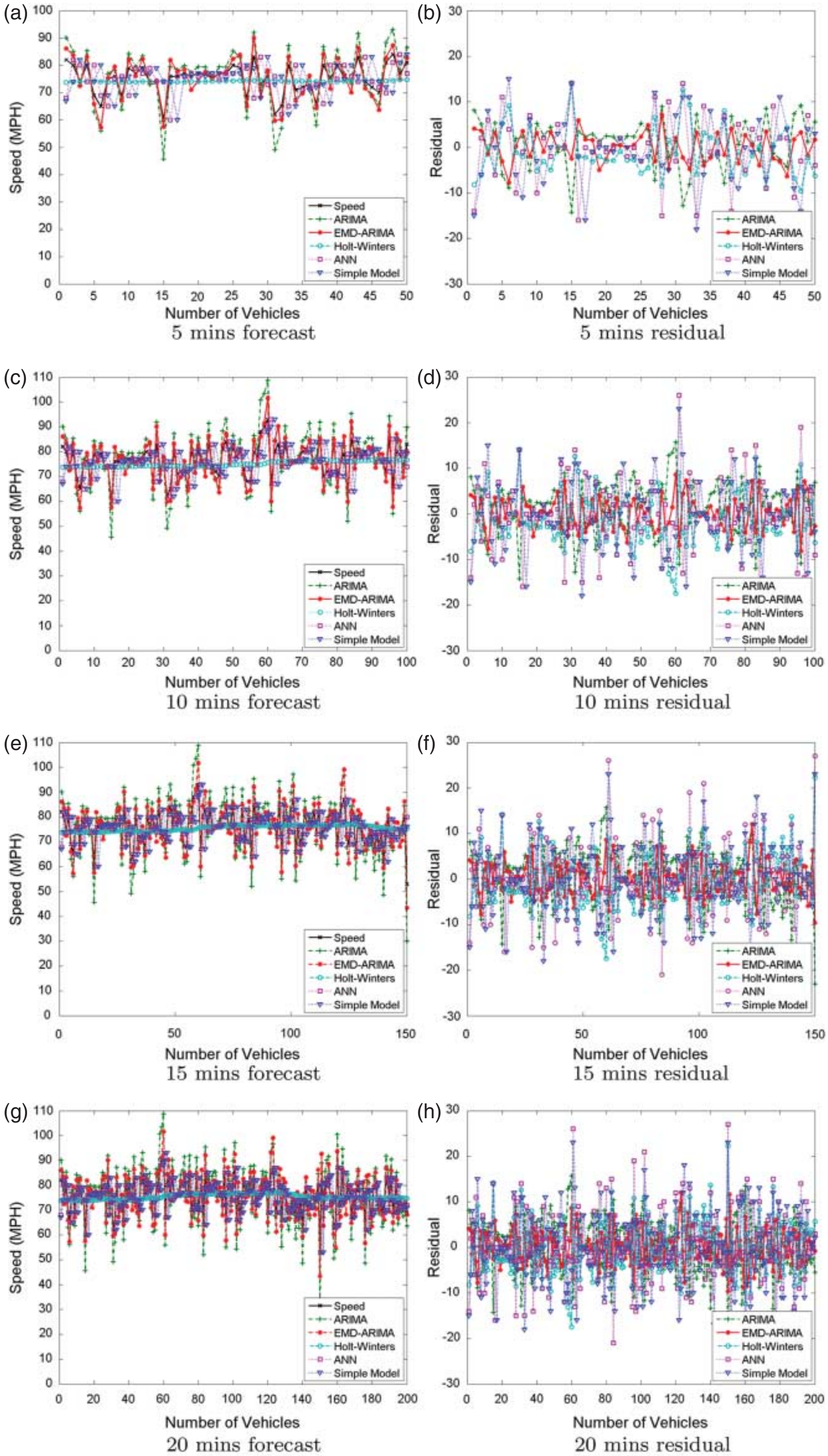

Figure 8. The short-term speed forecasting results on freeway off-ramps for 5, 10, 15, and 20-min horizons and its corresponding residuals for work zone scenario.

After we choose the start values, we can estimate the smoothing parameters by minimising the sum of the squared errors and finally get the forecasting results. Interested readers are referred to Chatfield (1978) for more details.

5. Analysis of results

In this section, we present a number of case studies on short-term traffic speed forecasting using the proposed EMD–ARIMA modelling framework for freeway basic segments, freeway on-ramp, and off-ramp, freeway work zone scenarios including mixed traffic flow and vehicle-type scenarios. For each individual scenario, the results generated from the EMD–ARIMA framework will be compared against a naive model, the Holt–Winters model, the traditional ARIMA model (ARIMA), and the ANN model to show its performance for different scenarios.

5.1. *Performance measures*

The mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) are calculated for each case to evaluate the accuracy of each model

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |x_i - \hat{x}_i|, \tag{15}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}, \tag{16}$$

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

Table 1. Analysis of model performance for 5, 10, 15, and 20-min forecasting horizon.

	5 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	2.84	4.36	2.70	5.74	5.98
RMSE (mph)	3.34	5.43	3.70	7.50	7.67
MAPE (%)	3.82	5.94	9.12	7.75	8.15
	10 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	3.04	5.05	4.37	6.20	6.34
RMSE (mph)	3.87	6.33	5.46	8.61	7.98
MAPE (%)	3.96	6.47	5.95	8.17	8.26
	15 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	3.49	4.60	4.94	6.04	6.20
RMSE (mph)	4.22	6.25	6.31	8.03	8.30
MAPE (%)	4.72	6.50	6.33	8.52	8.57
	20 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	3.16	4.72	4.69	6.08	6.41
RMSE (mph)	3.91	5.97	5.96	8.23	8.21
MAPE (%)	4.23	6.39	6.35	8.24	8.63

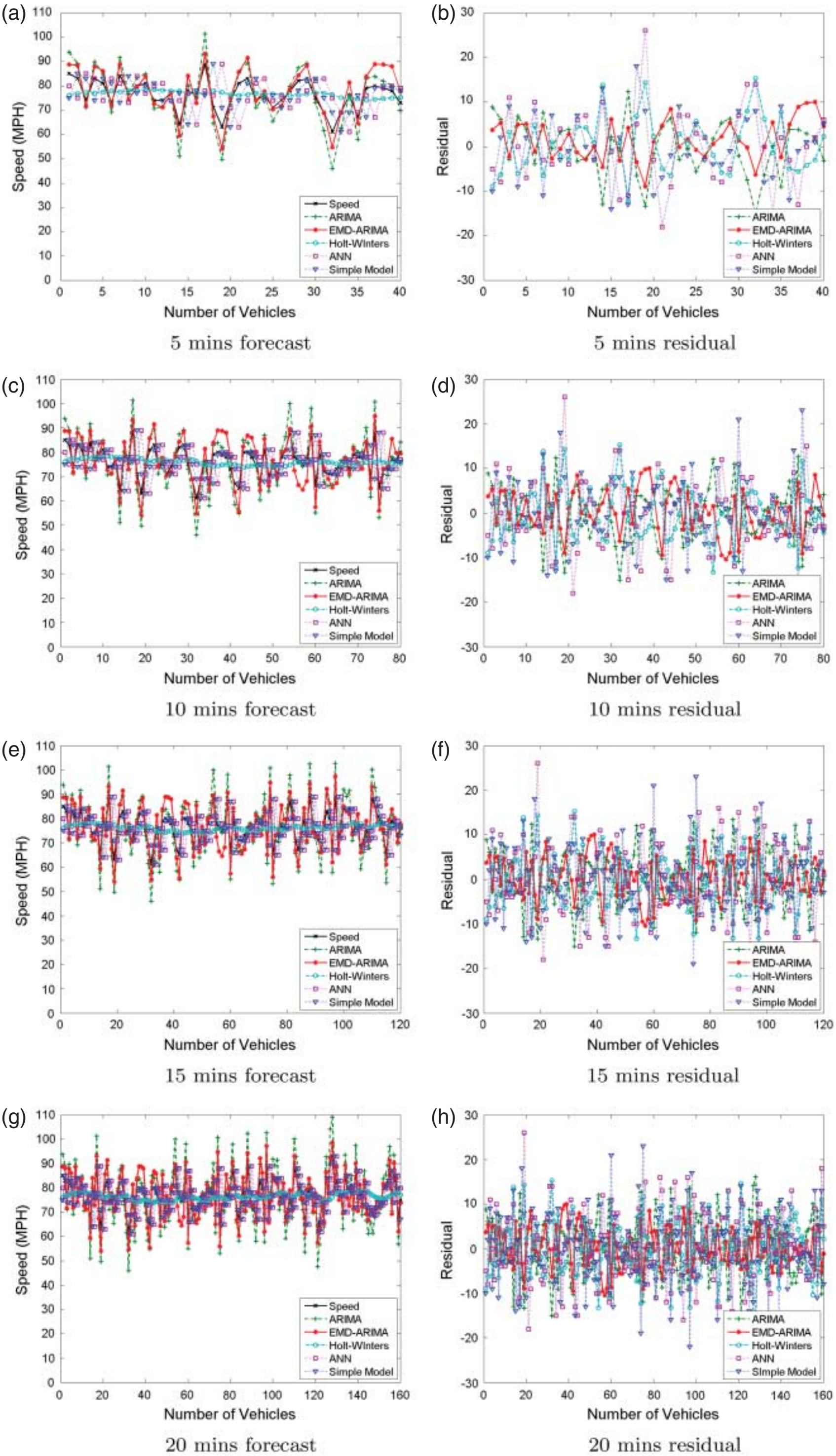

Figure 9. The short-term speed forecasting results on freeway off-ramps for 5, 10, 15, and 20-min horizons and its corresponding residuals for work zone scenario (cars).

in which x_i is the observation at the ith time interval, $\hat{x}_i$ is the predicted value at the ith time interval, and N is the number of observations.

The Holt–Winters model, the traditional ARIMA, the ANN model, and a naive model are used as the benchmarking forecasting methods to justify and evaluate the performance of the proposed EMD–ARIMA model. Four distinct test scenarios were built in this result section: (1) mixed traffic in a freeway work zone; (2) vehicle-type specific in a work zone; (3) freeway on-ramp; and (4) freeway off-ramp.

5.2. *Case study 1: mixed flow in freeway work zones*

The data were collected from a work zone segment on Nov 1st for 12 h (6:00 am–6:00 pm) in 2005 during which there were 7129 speed data in total. The forecast horizon is divided into four categories: 5, 10, 15, and 20 min, corresponding to 50, 100, 150 and 200 speed data. The forecasting results are shown in Figure 8. Table 1 presents the numerical comparison between the proposed EMD–ARIMA modelling framework and benchmarking forecasting methods (ARIMA, Holt–Winters, ANN and the simple model) in terms of MAE, MAPE, and RMSE.

The numerical comparison using the three proposed performance measures is referred in Table 1. Both results show that (1) the performance of EMD–ARIMA model is superior to those of the ARIMA model, the Holt–Winters model, the ANN model and the simple model, especially for the extreme values; (2) the performance of the Holt–Winters model is the second best, but it is close to the ARIMA model; (3) the EMD part of the hybrid model improves the performance of the traditional ARIMA model significantly: the MAPE of EMD–ARIMA model for the 5, 10, 15, 20 min forecasting horizons are 3.82%, 3.96%, 4.72%, 4.23%, respectively, compared to the MAPE of ARIMA model: 5.94%, 6.47%, 6.50%, 6.39%. On average, the EMD–ARIMA model is 2.14% superior to the traditional ARIMA model; (4) compared to the simple model which acts as the prediction basis, the MAPE of EMD–ARIMA model for the 5, 10, 15, 20 min forecasting horizons are 3.82%, 3.96%, 4.72%, 4.23%, respectively, and the MAPE of the simple model

Table 2. Analysis of model performance for 5, 10, 15, and 20-min forecasting horizon.

	5 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	3.74	4.87	5.10	7.05	6.35
RMSE (mph)	4.64	6.10	6.27	8.84	7.67
MAPE (%)	4.91	6.59	6.91	9.49	8.50
	10 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	3.99	4.72	4.92	6.58	6.26
RMSE (mph)	4.92	5.96	6.09	8.11	8.11
MAPE (%)	5.31	6.35	6.61	8.81	8.38
	15 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	3.69	4.80	5.01	6.98	6.35
RMSE (mph)	4.58	6.07	6.22	8.46	8.12
MAPE (%)	4.88	6.38	6.66	9.27	8.43
	20 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	3.43	4.97	5.18	6.83	6.37
RMSE (mph)	4.27	6.25	6.42	8.26	8.06
MAPE (%)	4.51	6.58	6.85	9.04	8.41

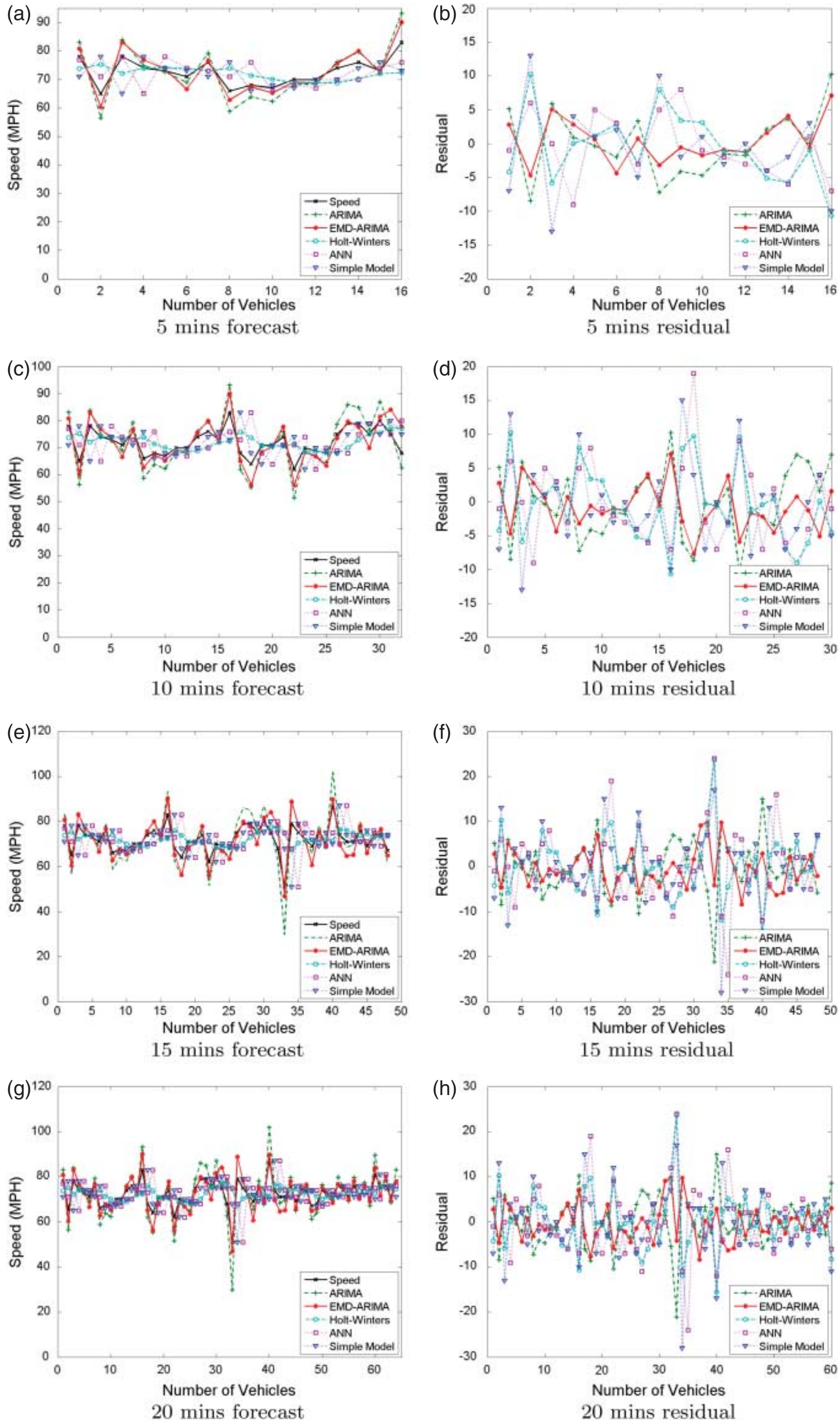

Figure 10. The short-term speed forecasting results on freeway off-ramps for 5, 10, 15, and 20-min horizons and its corresponding residuals for work zone scenario (trucks).

are 8.15%, 8.26%, 8.57%, 8.63%. The EMD–ARIMA model on average is 4.22% superior to the simple model; (5) compared to other models, the EMD–ARIMA model produces smaller prediction errors consistently regardless of the single- or multi-step short-term forecasting.

5.3. *Case study 2: vehicle-type specific in freeway work zones*

5.3.1. *The prediction of cars's speeds*

Figure 9 and Table 2 indicate that (1) the result of the speed prediction for the cars in the work zone is consistent with the result of mixed-traffic speed prediction in the work zone (the performance of the EMD–ARIMA model is the best, followed by the Holt–Winters model, the ARIMA model, the ANN model and the simple model); (2) the performance of the ARIMA model and Holt–Winters model is quite close to each other (the difference is less than 0.4%); (3) the performance of the EMD–ARIMA model in this scenario is not as good as that of the EMD–ARIMA model in the mixed-traffic work zone scenario, especially at the forecasting horizons of 5 and 10 min, indicating that if we treat the mixed-traffic flow as a whole to estimate the forecast of cars, we will probably overestimate the accuracy of our forecast, which is important if we want to implement the forecast results to ITS applications like ATIS; (4) lastly, compared to other models, the EMD–ARIMA model produces smaller prediction errors consistently regardless of the single- or multi-step short-term forecasting in this scenario.

5.3.2. *The prediction of trucks' speeds*

We can conclude from Figure 10 and Table 3 that (1) the result of the speed prediction for the trucks in work zone is again consistent with the result of mixed-traffic speed prediction in the work zone (the performance of the EMD–ARIMA model is the best, followed by the Holt–Winters model, the ARIMA model, the ANN model and the simple model); (2) the performance of the EMD–ARIMA model for trucks in the work zone is not as good as that of the EMD–ARIMA model in the mixed-traffic work zone scenario at the forecasting horizons

Table 3. Analysis of model performance for 5, 10, 15, and 20-min forecasting horizon.

	5 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	2.63	3.85	4.17	4.03	5.06
RMSE (mph)	3.25	4.80	5.23	4.78	6.52
MAPE (%)	3.59	5.31	5.76	5.54	6.86
	10 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	3.22	4.00	4.33	5.03	5.09
RMSE (mph)	4.08	4.97	5.52	6.38	6.53
MAPE (%)	4.52	5.61	6.07	7.14	7.14
	15 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	3.34	4.33	4.82	6.10	5.95
RMSE (mph)	4.22	5.91	6.61	8.16	8.09
MAPE (%)	4.71	6.21	6.91	8.76	8.37
	20 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	2.91	3.89	4.31	5.45	5.50
RMSE (mph)	3.76	5.33	5.94	7.29	7.38
MAPE (%)	4.08	5.52	6.12	7.75	7.65

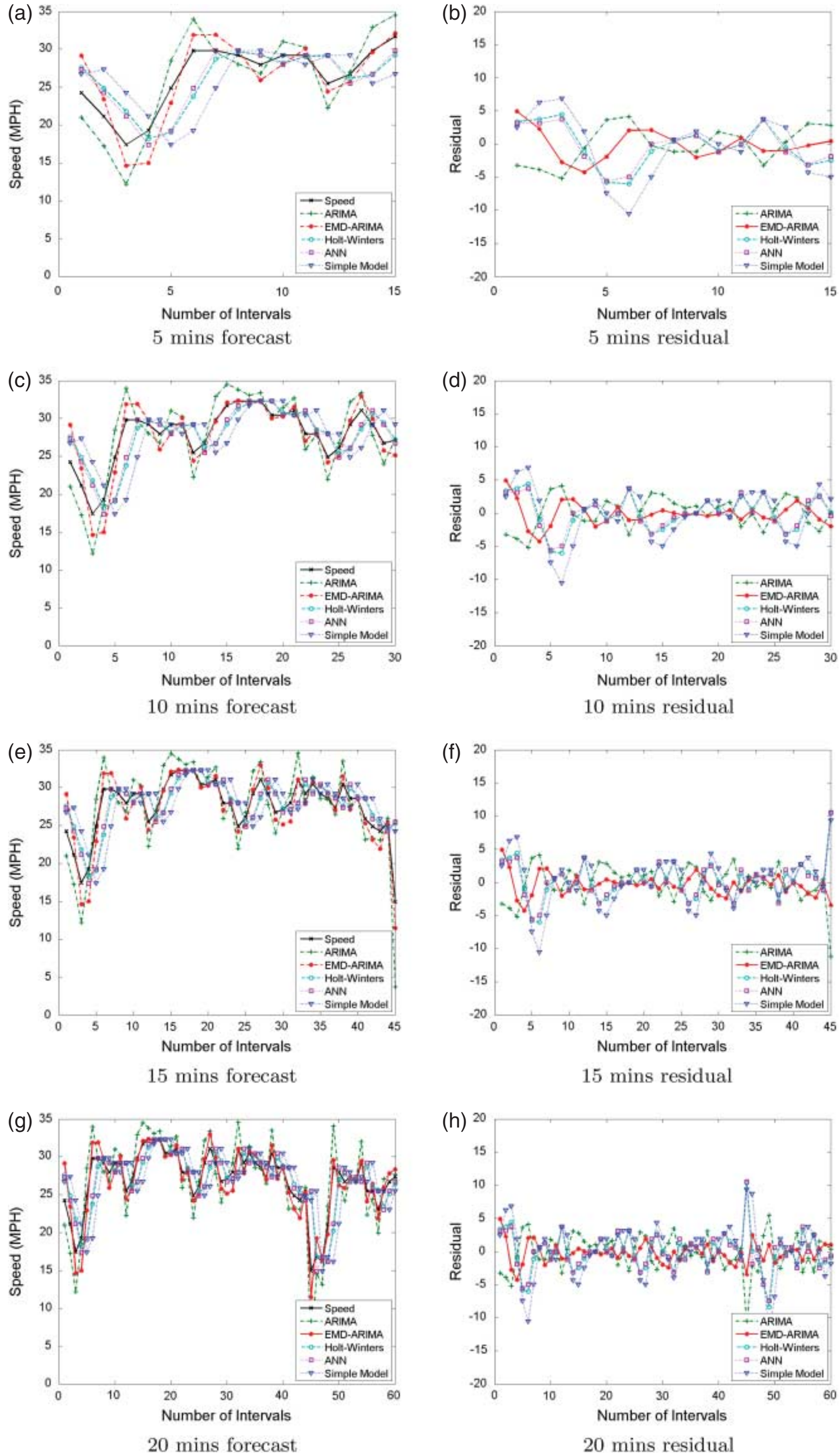

Figure 11. The short-term speed forecasting results on freeway on-ramps for 5, 10, 15, and 20-min horizons and its corresponding residuals.

of 5 and 10 min, indicating that if we treat the mixed traffic flow as a whole to estimate the forecast of trucks, we will probably overestimate the accuracy of our forecast especially at very short forecasting horizons; (3) an interesting finding is that if we compare the forecasting results of trucks and cars, we can see that the forecasting results of trucks are superior to the results of cars for all the forecasting horizons. We believe that this is associated with the interactions of the cars and trucks in the mixed-traffic, this interesting finding is worth further investigation; (4) lastly, compared to other models, the EMD–ARIMA model produces smaller prediction errors consistently regardless of single- or multi-step short-term forecasting in this scenario.

5.4. *Case study 3: on-ramp*

The speed prediction for on-ramps is conducted using a similar procedure for freeway basic segments. The testing of an on-ramp scenario is largely motivated by the fact that the on-ramp traffic flow is typically in an acceleration mode to join the basic freeway segment traffic speed which is different from either freeway or off-ramp scenarios. The results are presented in Figure 11 and Table 4.

Figure 11 and Table 4 indicate that (1) the result of the on-ramp speed prediction is consistent with the result of the work zone traffic prediction (the performance of the EMD–ARIMA model is the best, followed by the Holt–Winters model, the ARIMA model and the ANN model); (2) due to the larger variability of the original on-ramp speed series, the performance of all the models in this scenario (on-ramp) is inferior to the performance in the freeway basic segment scenario; (3) the performance of the EMD–ARIMA model in this scenario is a lot more superior to those of the ARIMA model, the Holt–Winters model and ANN model, indicating that, under the condition of larger speed variability, the proposed EMD–ARIMA model demonstrates its robustness; (4) compared with other models, the EMD–ARIMA model produces smaller prediction errors consistently regardless of the single- or multi-step short-term forecasting.

Table 4. Analysis of model performance for 5, 10, 15, and 20-min forecasting horizon.

	5 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	1.84	2.38	2.56	7.13	3.97
RMSE (mph)	2.26	2.81	3.17	8.13	4.87
MAPE (%)	7.80	9.70	10.33	16.13	15.86
			10 min forecasting results		
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	1.28	1.91	1.99	6.56	3.11
RMSE (mph)	1.73	2.33	2.59	7.50	3.93
MAPE (%)	5.18	7.34	7.69	14.12	11.87
			15 min forecasting results		
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	1.23	1.93	2.01	6.014	2.82
RMSE (mph)	1.65	2.70	2.78	7.08	3.68
MAPE (%)	5.07	8.03	8.27	13.43	11.27
			20 min forecasting results		
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	1.16	1.95	2.01	6.32	2.96
RMSE (mph)	1.54	2.64	2.82	7.39	3.98
MAPE (%)	4.83	8.15	8.19	14.25	12.03

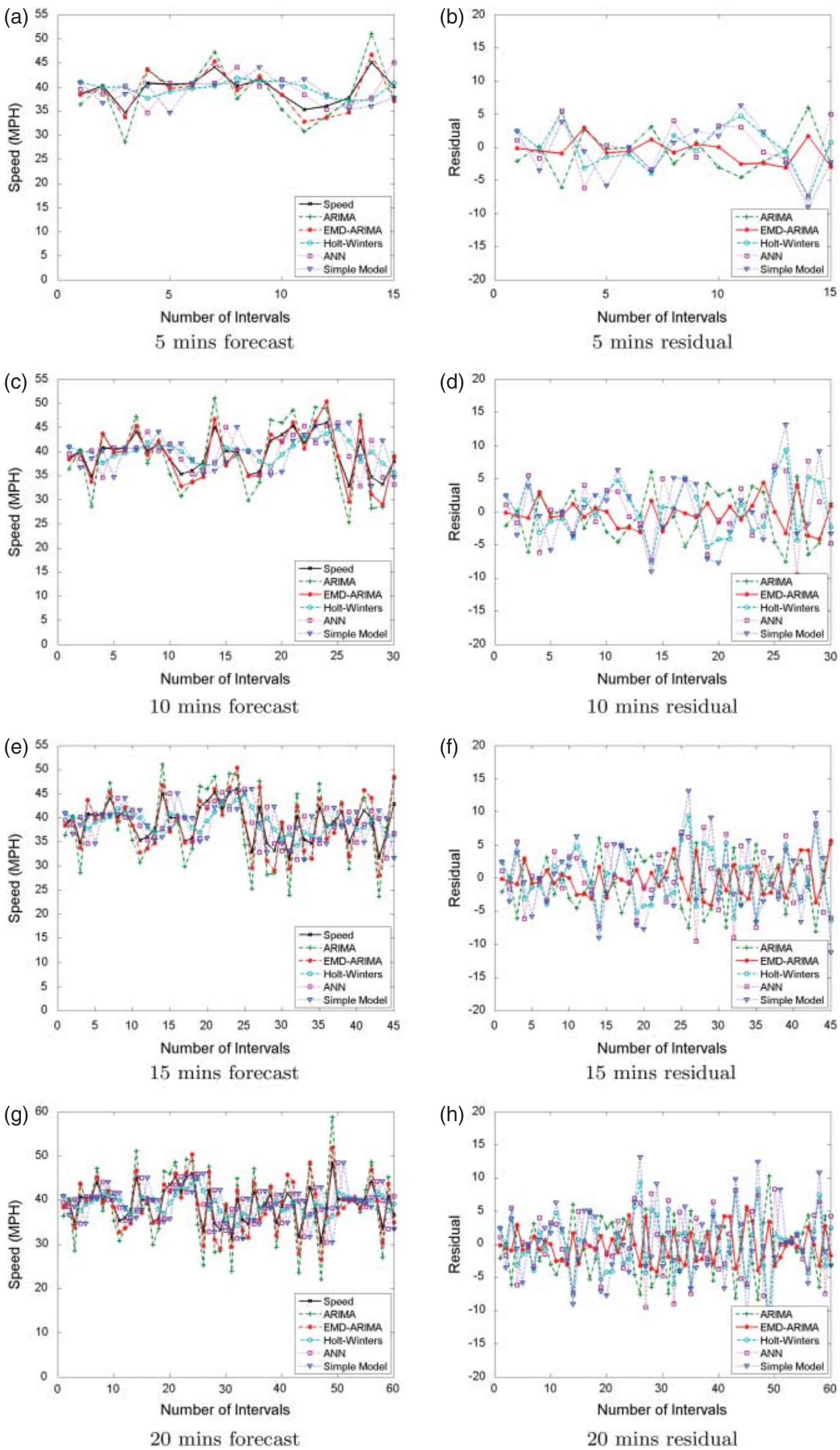

Figure 12. The short-term speed forecasting results on freeway off-ramps for 5, 10, 15, and 20-min horizons and its corresponding residuals.

Table 5. Analysis of model performance for 5, 10, 15, and 20-min forecasting horizon.

	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
	5 min forecasting results				
MAE (mph)	1.96	2.56	2.70	4.02	3.14
RMSE (mph)	2.38	3.48	3.70	4.75	3.90
MAPE (%)	6.68	8.65	9.12	13.31	7.94
	10 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	1.99	2.63	2.74	3.64	4.08
RMSE (mph)	2.38	3.39	3.54	4.37	5.03
MAPE (%)	7.29	9.42	9.76	12.90	10.59
	15 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	1.96	2.52	2.61	4.34	4.25
RMSE (mph)	2.34	7.19	3.25	7.08	5.14
MAPE (%)	7.19	9.04	9.31	15.41	11.13
	20 min forecasting results				
	EMD–ARIMA	ARIMA	Holt–Winters	ANN	Simple model
MAE (mph)	2.09	2.77	2.91	4.08	4.30
RMSE (mph)	2.49	3.48	3.66	4.90	5.64
MAPE (%)	7.60	10.03	10.47	14.57	11.24

5.5. *Case study 4: off-ramp*

Contrary to the acceleration mode at on-ramps, the vehicles are on average decelerating to negotiate the lower speed limit on off-ramps. The results are presented in Figure 12 and Table 5. It can be seen from Figures 12 and Table 5 that the result of the off-ramp speed prediction is rather consistent with the result of the basic freeway segment and on-ramp speed prediction: the performance of the EMD–ARIMA model is the best, followed by the Holt–Winters model, the ARIMA model, and the ANN model. This consistency is essential to show that the proposed hybrid EMD–ARIMA model is superior to the individual models in all four scenarios.

6. Conclusions and future work

The major contribution of this paper is the proposed hybrid EMD–ARIMA modelling framework which has been implemented for short-term freeway traffic speed prediction in four distinct scenarios: mixed traffic in work zone, vehicle-type specific prediction in a work zone, on-ramp, and off-ramp. One direct application of this modelling framework is to estimate link travel time from which the reliability analysis can be further conducted. The primary component of this hybrid model is an EMD, a recently developed method for non-stationary and nonlinear time-series analysis combined with an ARIMA. Therefore, this method is called EMD–ARIMA. This model delicately avoided the 'black-box' nature of other hybrid forecasting models such as EMD–ANN which has been applied to wind speed, passenger flow, and traffic speed prediction in recent literature.

The empirical data used in this effort were collected from Georgia State route 400 which is a heavily travelled corridor entering and exiting the city of Atlanta over a year's observation which includes a large spectrum of traffic conditions ranging from free flow to congestion and a freeway work zone on interstate I91 in Springfield, MA. From the four distinct short-term speed prediction scenarios including mixed and vehicle-type specific traffic prediction in work zone, on-ramp, and off-ramp, the results have shown that the hybrid EMD–ARIMA modelling framework is implementable and capable of generating accurate short-term traffic speed prediction from 5 to 20 min ahead. The hybrid EMD–ARIMA model was compared against the conventional prediction methods such as Holt–Winters, ARIMA, ANN, and a simple model subsequently. It was found that the proposed hybrid EMD–ARIMA modelling framework produced superior prediction results to the conventional methods. Numerical calculation of the prediction errors further verified its superiority over the traditional methods. From result analysis of the work zone scenarios, we found that if we use the mixed traffic speed forecast to estimate the forecast of trucks and cars, we will most probably overestimate the accuracy of our forecast especially at very short forecasting horizons; another interesting finding is that if we compare the forecasting results of trucks and cars, we can see that the forecasting results of trucks are superior to the results of cars for all the forecasting horizons. We believe this is due to the interactions of cars and trucks and we will further look into this issue.

The advantages of the proposed hybrid EMD–ARIMA method can be summarised as follows compared with the conventional methods.

- It is particularly suitable for traffic speed predictions because traffic speed series are non-stationary and nonlinear.
- The basis of the hybrid EMD–ARIMA model is derived from empirical data, therefore it does not require prior assumptions of the empirical data features (i.e. linear or nonlinear) while other methods such as Fourier analysis typically requires a sine or cosine basis function.
- The model is adaptive for varying traffic conditions (i.e. both free flow and congestion, stationary and non-stationary traffic conditions).

To the best knowledge of the authors, this research represents a leading effort of applying the hybrid EMD–ARIMA model to traffic-related predictions, acknowledging the prior application of this method in peak load demand forecasting by Okolobah and Ismail (2013). This method can be also applied to other traffic predictions such as passenger flow on both freeway and arterial roads. In addition, it is interesting to investigate which model is more reliable in the hybrid model families (EMD–ARIMA, EMD–ANN, EMD–SVM, etc.). There are more unknowns than what is known regarding the hybrid forecasting models, that is, the computation time and affordability of different methods.

Acknowledgements

The authors would like to thank the two anonymous reviewers' detailed and thoughtful comments which significantly helped the authors to improve the quality of the paper.

This article is an expansion of the TRB conference paper (Wang, Liu, Qian, Wei, and Dong 2014) and features following the new materials. (1) The authors have included an entire new section to review prior-related work to illustrate the importance and difficulties of short-term forecasting problem and identify the novelty of our work. (2) A more detailed and rigourous description for the proposed hybrid EMD–ARIMA forecasting framework and other alternative models compared in the experiments have been provided, which are at the same time more straightforward and easier to follow. (3) The experimental evaluations have been significantly expanded by adding three new scenarios featuring short-term traffic speed forecasting in freeway work zones for both mixed traffic flow and vehicle-type specific (i.e. car and truck) scenarios. A simple historical average (simple model) is added to the different case studies to provide a baseline prediction for the evaluation of other alternative models. Robustness of the model has been tested to demonstrate the model's applicability in a practical scenario.

References

Abdulhai, B., H. Porwal, and W. Recker. 2002. “Short-Term Traffic Flow Prediction Using Neuron-Genetic Algorithm.” *ITS Journal, Taylor and Francis* 7 (7): 3–41.

Administration, Federal Highway. 2003. “Manual on Uniform Traffic Control Devices for Highways and Streets Part 6: Temporary Traffic Control.” Technical report. US Dept. of Transportation, Washington, DC.

Bao, Y., T. Xiong, and Z. Hu. 2012. “Forecasting Air Passenger Traffic by Support Vector Machines with Ensemble Empirical Mode Decomposition and Slop-Based Method.” *Discrete Dynamics in Nature and Society* 2012, 12 (Article ID 431512). doi:10.1155/2012/431512.

Box, G. E. P., and G. M. Jenkins. 1976. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.

Castro-Neto, M., Y.-S. Jeong, M.-K. Jeong, and L. D. Han. 2009. “Online-svr for Short-Term Traffic Flow Prediction Under Typical and Atypical Traffic Conditions.” *Expert Systems with Applications* 36 (3): 6164–6173.

Chang, G.-L., and C.-C. Su. 1995. “Predicting Intersection Queue With Neural Network Models.” *Transportation Research Part C: Emerging Technologies* 3 (3): 175–191.

Chatfield, C. 1978. “The Holt–Winters Forecasting Procedure.” *Applied Statistics* 27 (3): 264–279.

Chen, B.-Y., W. H. K. Lam, Q. Li, A. Sumalee, and K. Yan. 2013. “Shortest Path Finding Problem in Stochastic Time-Dependent Road Networks with Stochastic First-In-First-Out Property.” *IEEE Transactions on Intelligent Transportation Systems* 14 (4): 1907–1917.

Chiou, J.-M., Y.-C. Zhang, W.-H. Chen, and C.-W. Chang. 2014. “A Functional Data Approach to Missing Value Imputation and Outlier Detection for Traffic Flow Data.” *Transportmetrica B: Transport Dynamics* 2 (2): 106–129.

Dougherty, M. S., and H. C. Kirby. 1998. “The Use of Neural Networks to Recognize and Predict Traffic Congestion.” *Traffic Engineering Control* 34 (6): 311–314.

Giebel, G., and G. Kariniotakis. 2009. “Best Practice in Short-Term Forecasting: A Users Guide.” Technical report. Report for the POW’WOW Project.

Hamad, K., M. T. Shourijeh, E. Lee, and A. Faghri. 2009. “Near-Term Travel Speed Prediction Utilizing Hilbert-Huang Transform.” *Computer-Aided Civil and Infrastructure Engineering* 24: 551–576.

Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu. 1998. “The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis.” *Proceedings of the Royal Society of London. Series A* 454: 903–995.

Innamaa, S. 2005. “Short-term Prediction of Travel Time Using Neural Networks on an Interurban Highway.” *Transportation* 32 (6): 649–669.

Jiang, X., and H. Adeli. 2005. “Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting.” *ASCE Journal of Transportation Engineering* 131 (10): 771–779.

Kamarianakis, Y., and P. Prastacos. 2003. “Forecasting Traffic Flow Conditions in An Urban Network: Comparison of Multivariate and Univariate Approaches.” *Transportation Research Record* 1857: 74–84.

Karlaftis, M. G., and E. I. Vlahogianni. 2009. “Memory Properties and Fractional Integration in Transportation Time-Series.” *Transportation Research Part C: Emerging Technologies* 17 (4): 444–453.

Koutroumanidis, T., K. Ioannou, and G. Arabatzis. 2009. “Predicting Fuelwood Prices in Greece with the Use of ARIMA Models, Artificial Neural Networks and a Hybrid ARIMA-ANN Model.” *Energy Policy* 37: 3627–3634.

Kumar, M., and M. Thenmozhi. 2012. “A Hybrid ARIMA-EGARCH and Artificial Neural Network Model in Stock Market Forecasting: Evidence for India and the USA.” *International Journal of Business and Emerging Markets* 4 (2): 160–178.

Lee, S., and D. B. Fambro. 1999. “Application of Subset Autoregressive Integrated Moving Average Model for Short-Term Freeway Traffic Volume Forecasting.” *Journal of Transportation Research Record* 1678: 179–188.

Liu, H., C. Chen, H.-Q. Tian, and Y.-F. Li. 2012. “A Hybrid Model for Wind Speed Prediction Using Empirical Mode Decomposition and Artificial Neural Networks.” *Renewable Energy* 48: 545–556.

Moorthy, C., and B. Ratcliffe. 1988. “Transportation Planning and Technology.” *Short Term Traffic Forecasting Using Time Series Methods* 12 (1): 45–56.

Mori, U., A. Mendiburu, M. çlvarez, and J. A. Lozano. 2014. "A Review of Travel Time Estimation and Forecasting for Advanced Traveller Information Systems." *Transportmetrica A: Transport Science* 11 (2): 1–39.

Ngo, L. B., A. Apon, and D. Hoffman. 2012. "An Empirical Study on Forecasting using Decomposed Arrival Data of an Enterprise Computing System." IEEE ninth international conference on information technology: new generations (ITNG), 756–763. http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber = 6209082

Okolobah, V., and Z. Ismail. 2013. "Forecasting Peak Load Demand by ARIMA and EMD Method." *Archives Des Sciences Geneve* 66 (1): 2–15.

Okutani, I., and Y. J. Stephanedes. 1984. "Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory." *Transportation Research Part B: Methodological* 18B: 1–11.

Pai, P.-F., and C.-S. Lin. 2005. "A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting." *Omega* 33 (6): 497–505.

Pan, T. L., A. Sumalee, R.-X. Zhong, and N. Indra-Payoong. 2013. "Short-Term Traffic State Prediction Based on Temporal-spatial Correlation." *IEEE Transactions on Intelligent Transportation Systems* 14 (3): 1242–1254.

Papageorgiou, M., and Y. Wang. 2007. "Real-Time Freeway Traffic State Estimation Based on Extend Kalman Filter: A Case Study." *Transportation Science* 42 (2): 167–181.

Smith, B. L., and M. J. Demetsky. 1994. "Short-Term Traffic Flow Prediction: Neural Network Approach." *Transportation Research Record* 1453: 98–104.

Smith, B. L., B. M. Williams, and K. R. Oswald. 2002. "Comparison of Parametric and Non-Parametric Models for Traffic Flow Forecasting." *Transportation Research Part C: Emerging Technologies* 10 (4): 303–321.

Stathopoulos, A., and M. G. Karlaftis. 2003. "A Multivariate State Space Approach for Urban Traffic Flow Modeling and Prediction." *Transportation Research Part C: Emerging Technologies* 11 (2): 121–135.

Taylor, D. R., S. Muthiah, B. T. Kulakowski, K. M. Mahoney, and R. J. Porter. 2007. "Artificial Neural Network Speed Profile Model for Construction Work Zones on High-Speed Highways." *ASCE Journal of Transportation Engineering* 133 (3): 198–204.

Ullman, G. L., A. J. Holick, T. Scriba, and S. M. Turner. 2004. "Estimates of Work Zone Exposure on the National Highway System in 2001." *Transportation Research Record* 1877: 62–68.

Van Der Voort, M., M. Dougherty, and S. Watson. 1996. "Combining Kohonen Maps with Arima Time Series Models to Forecast Traffic Flow." *Transportation Research Part C: Emerging Technologies* 4 (12): 307–318.

Vlahogianni, E. I., J. C. Golias, and M. G. Karlaftis. 2004. "Short-term Traffic Forecasting: Overview of Objectives and Methods." *Transport Reviews* 24 (5): 533–557.

Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. 2005. "Optimized and Meta-Optimized Neural Networks for Short-Term Traffic Flow Prediction: A Genetic Approach." *Transportation Research Part C: Emerging Technologies* 13 (3): 211–234.

Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. 2006. "Statistical Methods for Detecting Nonlinearity and Non-Stationarity in Univariate Short-Term Time-Series of Traffic Volume." *Transportation Research Part C: Emerging Technologies* 14 (5): 351–367.

Vlahogianni, E. I., M. G. Karlaftis, and J. C. Golias. 2008. "Temporal Evolution of Short-Term Urban Traffic Flow: A Nonlinear Dynamics Approach." *Computer-Aided Civil and Infrastructure Engineering* 22 (5): 317–325.

Wang, Haizhong, Lu Liu, Zhen Qian, Heng Wei, and Shangjia Dong. 2014. "Empirical Mode Decomposition-Autoregressive Integrated Moving Average Hybrid Short-Term Traffic Speed Prediction Model." *Transportation Research Record: Journal of the Transportation Research Board* 2460: 66–76.

Watson, S. M., H. R. Kirby, and M. S. Dougherty. 1997. "Should We Use Neural Networks or Statistical Models for Short-Term Motorway Traffic Forecasting?" *International Journal of Forecasting* 13 (1): 43–50.

Wei, Y., and M.-C. Chen. 2012. "Forecasting the Short-Term Metro Passenger Flow with Empirical Mode Decomposition and Neural Networks." *Transportation Research Part C: Emerging Technologies* 21 (1): 148–162. http://www.sciencedirect.com/science/article/pii/S0968090X11000994.

Williams, B. M., P. K. Durvasula, and E. B. Donald. 1998. "Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models." *Transportation Research Record* 1644: 131–141.

Wunderlich, K. 2005. "Characterizing Nationwide Work-Zone Activity: Key Findings from Website Content Analyses, 2001–2003." Transportation Research Board 84th annual meeting. http://www.itsdocs.fhwa.dot.gov//JPODOCS/REPTS_TE//13793.html.

Xie, Y., Y. Zhang, and Z. Ye. 2007. "Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition." *Computer-Aided Civil and Infrastructure Engineering* 22 (5): 326–334.

Ye, Z., and Y. Zhang. 2008. "Short-Term Traffic Flow Forecasting Using Fuzzy Logic System Methods." *Journal of Intelligent Transportation Systems* 12 (3): 102–112.

Zhu, B., and Y. Wei. 2013. "Carbon Price Forecasting with a Novel Hybrid ARIMA and Least Squares Support Vector Machines Methodology." *Omega* 41: 517–524.