Empirical Mode Decomposition–Autoregressive Integrated Moving Average

Article *in* Transportation Research Record Journal of the Transportation Research Board · December 2014

DOI: 10.3141/2460-08

Haizhong Wang
Oregon State University

Lu Liu
Oregon State University

Sean Qian
Carnegie Mellon University

Heng Wei
University of Cincinnati

Empirical Mode Decomposition–Autoregressive Integrated Moving Average

Hybrid Short-Term Traffic Speed Prediction Model

Haizhong Wang, Lu Liu, Zhen (Sean) Qian, Heng Wei, and Shangjia Dong

Short-term freeway traffic speed prediction is essential to improving mobility and roadway safety. It has been a challenging and unresolved issue. Traffic speed prediction can be applied to enhance the intelligent freeway traffic management and control for applications such as operational and regulation planning. For example, with more reliable traffic speed prediction, the advanced traveler information system can provide travelers with predictive travel time information and optimal routing, which allows them to arrange their schedules accordingly. Moreover, traffic managers can use the predicted information to deploy various traffic management strategies to increase system efficiency. In this paper, a hybrid empirical mode decomposition (EMD) and autoregressive integrated moving average (ARIMA) (or EMD-ARIMA) approach was developed to predict the short-term traffic speed on freeways. In general, there were three stages in the hybrid EMD-ARIMA forecasting framework. The first was the EMD stage, which decomposed the freeway traffic speed time series data into a number of intrinsic mode function (IMF) components and a residue. The second stage was to find the appropriate ARIMA model for each IMF and residue and then make predictions on the basis of the appropriate ARIMA model. The third stage was to combine the prediction results of each IMF and residue to make the predictions. The experimental results indicated that the proposed hybrid EMD-ARIMA framework was capable of predicting short-term freeway traffic speed with high accuracy.

In a highly mobile society, travelers have an increasing desire and need for accurate and timely travel information to help them decide how to reach their destinations safely and efficiently. Emerging information and communication technologies play increasingly important roles in the modern transportation system while allowing information access across varying geographic scales and time frames. This multifunctionality has opened the door to a wide range of new opportunities (i.e., better prediction of future conditions) to improve transportation mobility, accessibility, and safety, while reducing emissions from travelers, steering transportation systems toward more sustainable and socially desirable directions. With this growing need, intelligent transportation systems have been pursued as the common technological approach to meet the demand. However, a reliable real-time traffic speed prediction is essential to intelligent transportation system management and control, especially for an advanced traveler information system and advanced traffic management system (*1*). With accurate traffic speed prediction, an advanced traveler information system can provide travelers with travel time information and optimal routing, allowing them to arrange their schedules accordingly. Moreover, traffic managers can use the prediction information to deploy various traffic management strategies through advanced traffic management system so as to increase system efficiency. In particular, short-term traffic speed prediction is of paramount significance to forecast future traffic conditions in minute intervals (i.e., 5 min or 10 min) to provide the traveling public with traffic information in a timely manner. Therefore, reliable short-term traffic speed prediction is the fundamental premise to the real-time intelligent transportation system.

Traditionally, researchers and practitioners have been using statistical methods [i.e., the autoregressive integrated moving average (ARIMA) model] and artificial intelligence (AI)–based methods [i.e., the artificial neural network (ANN)] in that regard. Both methods have achieved modest successes while bearing obvious limitations, that is, the linear assumption of the ARIMA model and the unknown mechanism of the ANNs method. For the high nonlinearities and irregularities of the traffic speed series, it is often contrary to the linear assumptions of statistical methods. In addition, AI-based methods such as ANN serve as a "black box," with their pattern structure hidden from users. The proposed hybrid empirical mode decomposition–ARIMA (EMD-ARIMA) is designed to overcome these limitations to some extent, by applying EMD to decompose the original series for insights. This method is able to deal with nonlinear and nonstationary series, and compared with AI-based methods, this method enjoys a more transparent and explicit structure with improved prediction accuracy.

The major contribution of this research is threefold: (*a*) a hybrid EMD-ARIMA model is proposed and applied to short-term traffic speed predictions in three distinct scenarios: basic freeway segment, on-ramp, and off-ramp; (*b*) the prediction results of the proposed model are evaluated through measures such as the mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE), and the results outperformed the traditional ARIMA models, the Holt–Winters model, and the ANN model; and (*c*) the experiments are conducted on basic freeway segments for different time horizons and the results are consistently superior to traditional models.

H. Wang, L. Liu, and S. Dong, School of Civil and Construction Engineering, Oregon State University, 101 Kearney Hall, 1491 SW Campus Way, Corvallis, OR 97331. Z. Qian, H. John Heinz III College, Carnegie Mellon University, 2102 C Hamburg Hall, 5000 Forbes Avenue, Pittsburgh, PA 15213. H. Wei, Department of Civil and Architectural Engineering and Construction Management, University of Cincinnati, 729 Rhodes Hall, 2901 Woodside Drive, Cincinnati, OH 45221-0071. Corresponding author: H. Wang, Haizhong.Wang@oregonstate.edu.

Transportation Research Record: Journal of the Transportation Research Board, No. 2460, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 66–76.
DOI: 10.3141/2460-08

The rest of this paper is organized as follows. A literature review on different models for short-term forecasting is presented in the next section, followed by a description of the framework of the hybrid EMD-ARIMA model and the detailed modeling steps of the EMD method and the ARIMA model. The empirical data and the study site of this research are then presented. The performance of the hybrid prediction results on the freeway basic segment for single and multiple-step forecasting horizon (i.e., 5 min, 10 min, 15 min, and 20 min) is shown. Last, the research is concluded and future research work is discussed.

LITERATURE REVIEW

During the past decades, researchers and practitioners have developed numerous prediction methods, which can be categorized as (*a*) statistical time series models, (*b*) AI-based models, and (*c*) hybrid forecasting models. Statistical time series models generally include a spectrum of time series models such as ARIMA models or Bayesian-based prediction mechanisms. The AI-based prediction methods include ANN, fuzzy logic methods, and support vector machine (SVM). Hybrid forecasting models typically combine two or more models to form synergetic prediction models. Some examples of hybrid models are EMD-ANN, EMD-SVM, or ensemble empirical mode decomposition (EEMD)-ANN. The literature about implementing these methods to various application domains abounds. The pros and cons of each method have been documented and understood in the literature. A summary of the available models is shown in Figure 1.

Next, a review and summary of the existing relevant studies in the three aforementioned categories will be provided. This review, however, is not intended to be comprehensive.

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 nonparametric) to relatively sophisticated methods such as the Holt–Winters forecasting procedure (*2*) and the Box–Jenkins methodology (*3*). In addition, Moorthy and Ratcliffe explored the application of time series analysis to produce short-term forecasts and investigated the performance of the forecasting models (*12*). Lee and Fambro investigated the use of the ARIMA model for short-term traffic volume forecasting, which showed that the use of a subset of ARIMA models would increase the accuracy of the short-term forecasting performance in time series models (*13*). Van Der Voort et al. introduced a hybrid short-term forecasting method known as the KARIMA method (*14*). They used the Kohonen self-organizing map as an initial classifier, which showed improvement over the single ARIMA model. Kamarianakis and Prastacos presented and discussed two different univariate (historical average and ARIMA) methods and two different multivariate methods [vector autoregressive moving average (VARMA) and space–time ARIMA (STARIMA)] and compared the performance of the four models (*15*). Their work showed comparable results for the ARIMA, VARMA, and STARIMA models. Williams et al. applied a seasonal ARIMA and exponential smoothing model to predict urban freeway traffic flow (*16*).

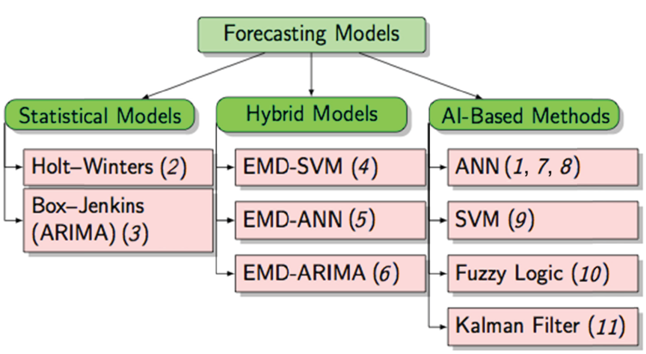

FIGURE 1 Summary of available forecasting models.

AI-Based Methods

In general, the aforementioned prediction methods 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 traditional prediction methods (*1, 4, 17*). Because of the inherent limitation, many AI models (i.e., ANNs, SVM, and Kalman filter) are used in short-term forecasting research. There have been many discussions on the use of the ANN for short-term predictions. Smith and Demetsky (*7*) and Dougherty and Kirby (*8*) were among the first researchers who successfully applied the neural network to short-term traffic flow forecasting such as traffic congestion prediction. Later, Jiang and Adeli (*18*) and Innamma et al. (*19*) used the neural network to predict short-term traffic flow performance and demonstrated the ability to project future traffic conditions. However, the neural network–based prediction mechanism has been criticized by its users for its inherent black box nature and hidden structure and the intensive demand of training data although the prediction accuracy seems appealing for practical implementation purposes. Watson et al. questioned the neural network approach and raised the question of whether statistical methods or neural networks should be used for short-term traffic flow forecasting (*20*). In this study, the authors pointed out that in the application of traffic series, the issue of the neural network's black box nature with its structure hidden from the user needed to be considered more deeply.

For other AI models, Castro-Neto et al. conducted online short-term traffic flow prediction through an SVM-based approach for normal and abnormal traffic conditions (*9*). Likewise, Xie et al. suggested the Kalman filter and discrete wavelet decomposition approach (*21*). Later, Ye and Zhang performed short-term traffic flow forecasting through a fuzzy logic system approach (*10*). Finally, Papageorgiou and Wang proposed an extended Kalman filter–based approach to estimate real-time freeway traffic states (*11*).

Hybrid Forecasting Models

The forecasting results of 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 programming and SVM are too sensitive in selecting parameters and the local minima. Overfitting is also an issue. To correct these flaws, some hybrid methods are developed to address the short-term traffic forecasting and have proved to be effective.

Researchers used the fusion model of EMD, ANN, genetic algorithm, and SVM to make predictions on air traffic flow (*4*), wind speed

(*5*), and metro passenger flow (*17*). Bao et al. proposed the hybrid EEMD-SVM approach to forecast air traffic passenger flow with the use of six airlines' empirical data (*4*). The results showed that the proposed hybrid modeling methodology outperforms the pure SVMs, Holt–Winters, and ARIMA. Abdulhai et al. trained the neural network with genetic algorithms to forecast short-term freeway traffic flow (*22*). Hamad et al. investigated the near-term travel speed forecasting by using a Hilbert–Huang transform, which is actually a hybrid model of EMD and backpropagation neural network (*1*). Ngo et al. used several forecasting techniques (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 (*23*). Results showed that EMD helps to improve forecasting results. Okolobah and Ismail proposed a method based on EMD and ARIMA to predict the peak load, and the result showed that the proposed model presented better forecasting accuracy than the traditional ARIMA method (*6*). The primary motivation of these hybrid methods is to fix the individual flaws and generate a synergetic way to accomplish the short-term prediction of wind speed, traffic flow, passenger flow, and so on.

HYBRID EMD-ARIMA MODELING FRAMEWORK

Inspired by the hybrid methods, this paper proposes a hybrid method of EMD-ARIMA for short-term traffic speed forecasting. In this methodology, the EMD technique is used to decompose the original traffic speed data into independent components. The main purpose of decomposition is twofold: (*a*) to decompose the original nonstationary traffic speed series into stationary subseries to facilitate the forecasting and (*b*) to differentiate the modes with different characters [i.e., intrinsic mode functions (IMFs)] contributing to the prediction accuracy. Those IMFs may imply insights for recognizing time series patterns of speeds. The combined procedure is to develop a consensus prediction on immediate historical data. With the use of the EMD techniques, the original traffic speed series with nonlinear and nonstationary features are decomposed into a finite and small number of IMFs. After these IMFs are obtained and each IMF is modeled by an ARIMA model, they can be forecast independently. Finally, all the predicted results are aggregated to generate a combined forecasting result. A detailed procedure of the EMD process is presented in Figure 2.

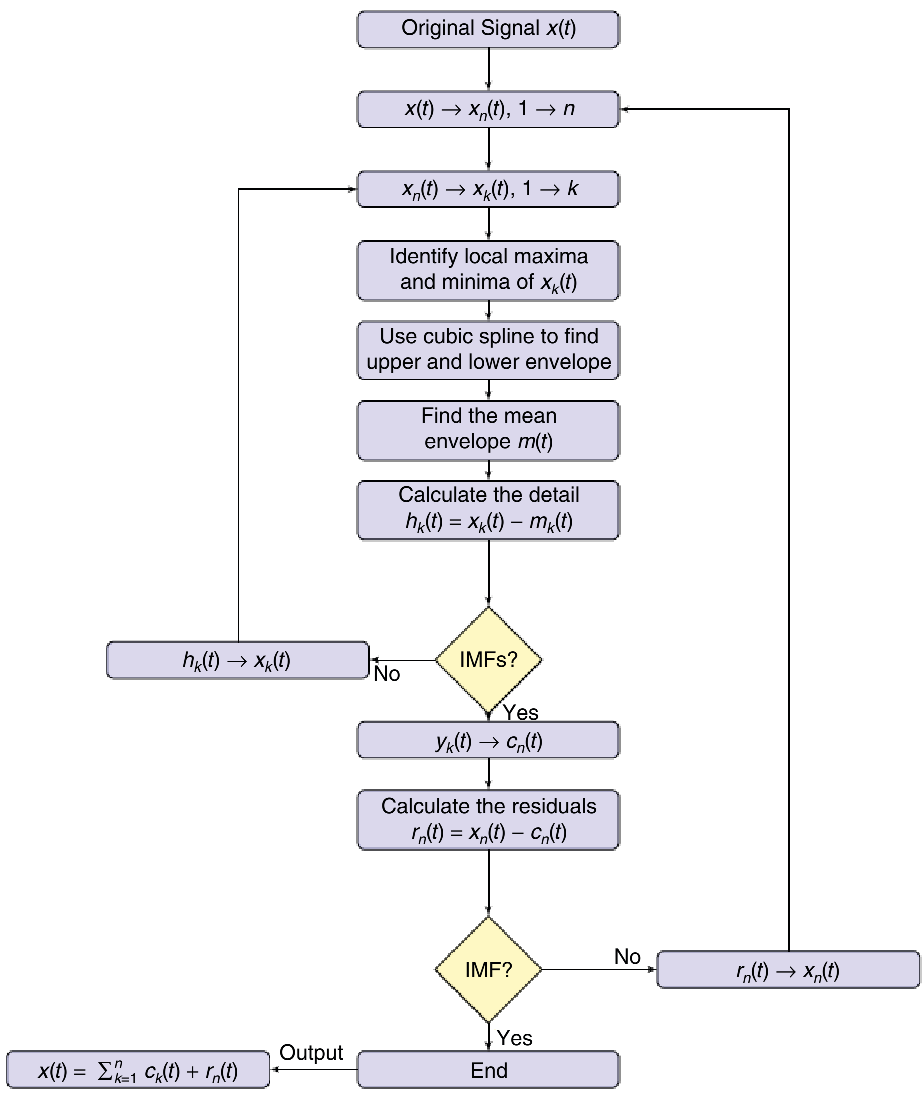

FIGURE 2 Flowchart of empirical mode decomposition algorithm.

The hybrid modeling framework through EMD-ARIMA includes two major steps: (*a*) EMD and (*b*) ARIMA. Traditionally, researchers have been using time–frequency analysis algorithms such as short-time Fourier transform or wavelet transform to decompose the original signal into a series of subseries. The major difference between wavelet transform and EMD is that wavelet transform uses the wavelet function (such as Harr, Morlet, and Daubechies) to decompose the original signal while EMD decomposes the signal through an iterative shifting process to extract the IMFs. EMD is particularly suitable for a nonlinear and nonstationary signal, while wavelet transform is more appropriate for a linear signal. The essence of the EMD is to decompose the original signal into the sum of a finite number of IMFs and the mean trend. The decomposition process of EMD is based on the local characteristic time slice to obtain the IMFs. This hybrid-modeling framework has been applied to predict the arrival data of an enterprise computing system (*23*) and peak load demand forecasting (*6*), with great success.

Stage 1. EMD

Following the pioneering work of Huang et al. on the Hilbert–Huang transform, the EMD has been widely used to decompose a signal into the IMF and use the Hilbert spectrum analysis to analyze nonstationary and nonlinear time series data (*24*). Empirical traffic speed series over time is a nonstationary and nonlinear time series. Because of the nonlinear and nonstationary nature of traffic speed series, by decomposing the time series speeds into basic components (IMF), more accurate predictions would be obtained.

Any original traffic volume or speed series defined by $x(t)$, $t = 1, 2, \ldots, T$, can be written in the following form after the empirical mode decomposition algorithm depicted in Figure 2 is applied:

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

where $r(t)$ 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:

1. 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.
2. 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 "shifting process" is adopted to extract the separate IMF components. The detailed procedures of the shifting process can be summarized as follows:

1. Identify all the local extrema (i.e., local maxima and minima) in the time series $x(t)$.
2. Interpolate all the local maxima by a cubic spline line to form the upper envelope $e_{\max}(t)$.
3. Repeat the procedure for the local minima to produce the lower envelope $e_{\min}(t)$.
4. Compute the mean envelope $m(t)$ from the upper (high) and lower envelope as follows: $m(t) = [e_{\max}(t) + e_{\min}(t)]/2$.
5. Extract the mean envelope from the original signal to obtain a new mean envelope as follows: $m(t) = x(t) - h(t)$.
6. Check whether $h(t)$ is an IMF: (*a*) if $h(t)$ is an IMF, then set $d(t) = h(t)$, and at the same time replace $x(t)$ with the residual $r(t) = x(t) - d(t)$ and (*b*) if $h(t)$ is not an IMF, replace $x(t)$ with $z(t)$ and then repeat the steps from 2 to 5 until the following stopping criterion is met in the iterative process.

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

Stage 2. ARIMA Model Design

ARIMA models are the most general category of the time series model, which is widely used in short-term forecasting. An ARIMA model is typically used when the time series is stationary or can be stationarized by transformations such as differencing. A stationary time series is one that has a constant mean, variance, and autocorrelation.

The different series appearing in the forecasting equation are called "autoregressive" terms; the forecasting errors are called "moving average" terms. The simple nonseasonal ARIMA is classified as an "ARIMA (p, d, q)" model, where p is the number of autoregressive terms, d is the number of nonseasonal differences, and q is the number of lagged forecasting errors. In an ARIMA model, the future value of a variable is supposed to be a linear combination of past values and past errors.

A general ARIMA model is defined by the following equation:

$$y(t) = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

where

$y(t)$ = actual value,
ε_t = random error at time t,
ϕ_i and ϕ_j = coefficients,
ϕ_p = parameters of autoregressive part,
θ_q = parameters of moving average part, and
p and q = autoregressive and moving average polynomials, respectively.

Basically, the model has three phases: model identification, parameter estimation, and diagnostic checking. The framework of the two-stage modeling is shown in Figure 3.

EXPERIMENTS

Data Description

The empirical data used in this research were collected through the Georgia intelligent transportation system program by the Georgia Department of Transportation on SR-400, which is the south–north freeway entering and exiting the city of Atlanta. The empirical data were collected over an approximately 20-mi-long corridor including 100 detectors in which 78 detectors were installed on the basic freeway segments and 22 deployed on the entrance and exit ramps. The geographic site and the study area and GA-400 are shown in Figure 4.

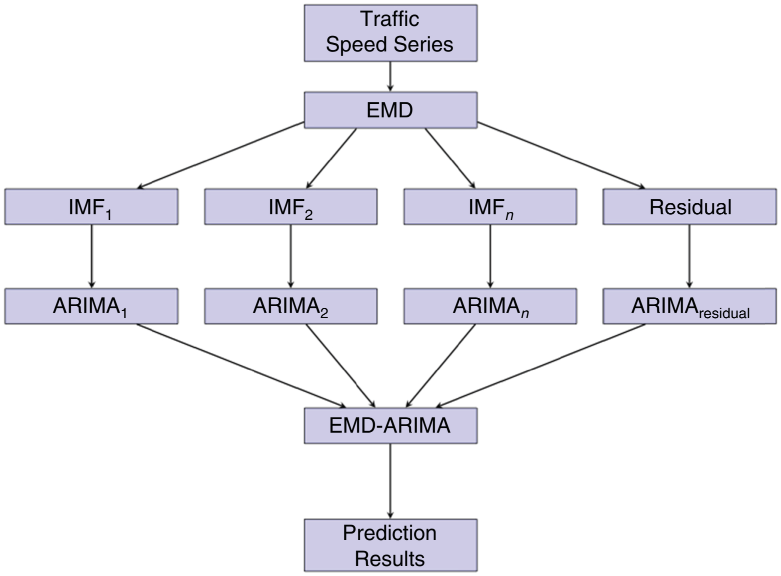

FIGURE 3 Hybrid EMD-ARIMA modeling framework.

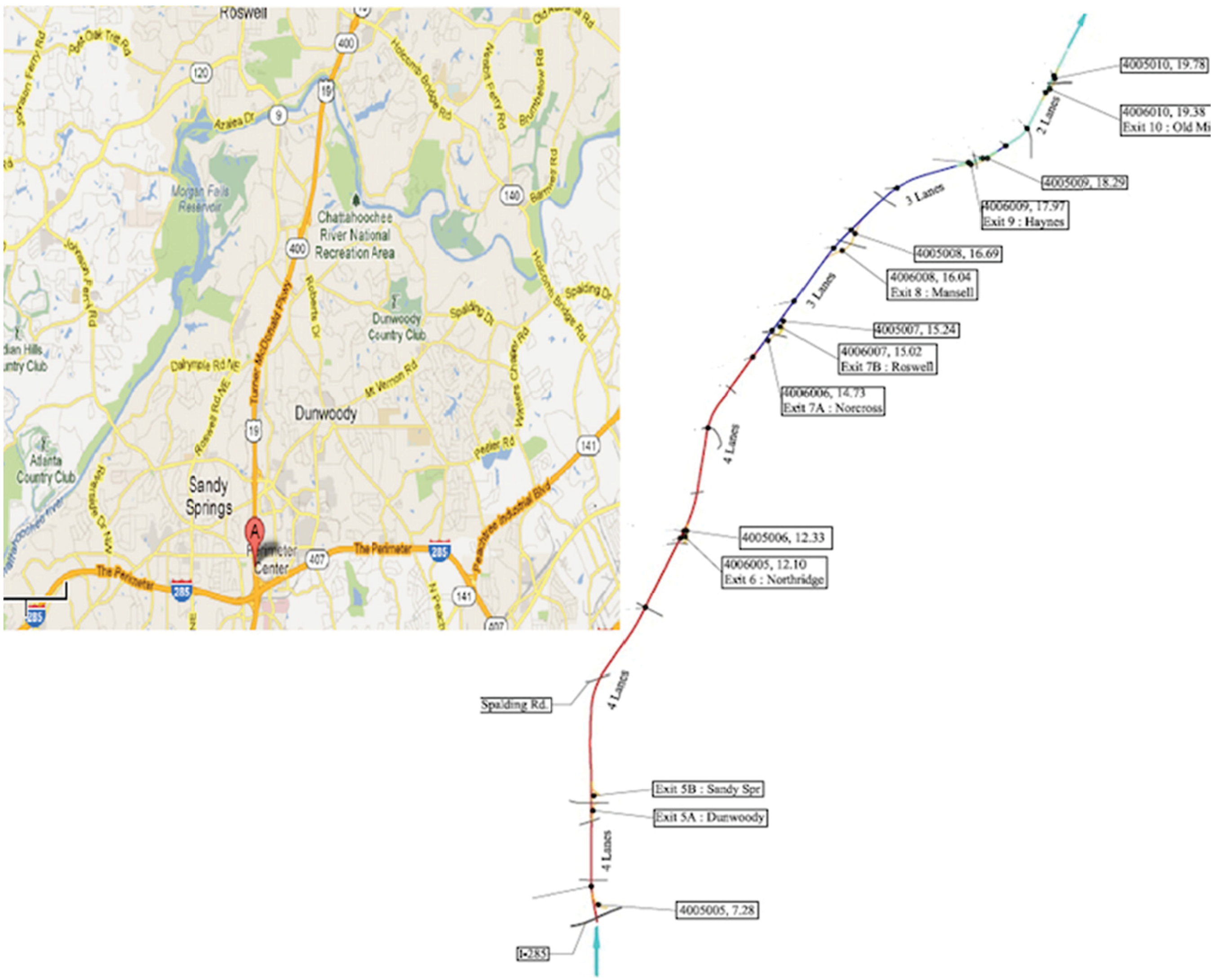

FIGURE 4 Selected study site from GA-400 northbound at Atlanta. (Source: Google Maps and hand sketches.)

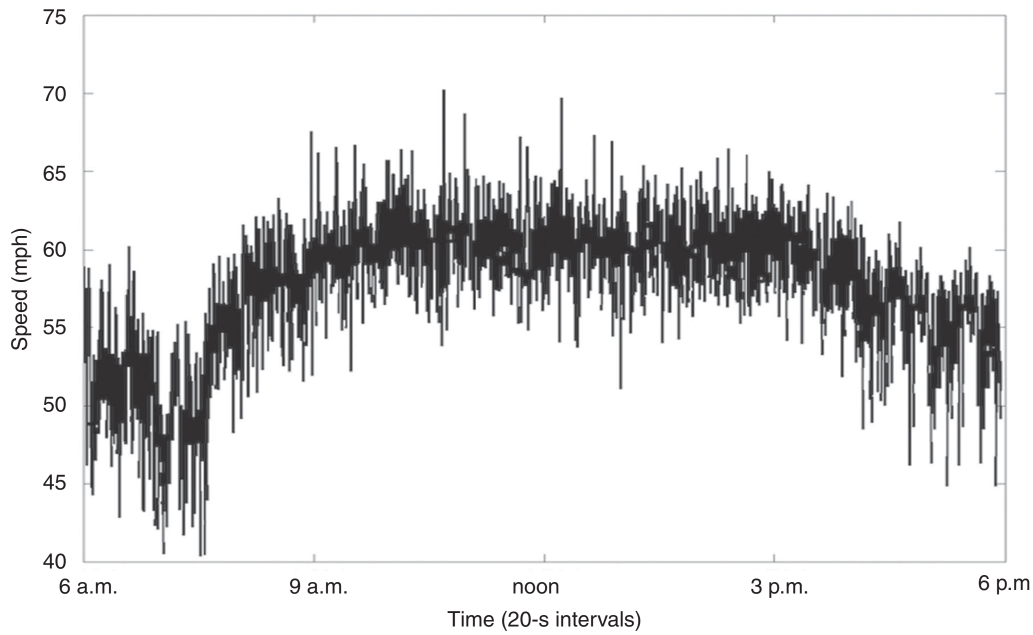

FIGURE 5 Traffic speed series over time from 1-week observation of freeway basic segment.

The data used in this experiment are taken from the basic freeway segment traffic speed data. As can be seen from Figure 5, the 20-s average speed for the morning peak hours (7:00 to 9:00 a.m.) were collected over a period of 5 weekdays. The data for the first 4 days were used as the training data set to build the hybrid EMD-ARIMA forecasting model, while the data for the 5th day were used as the empirical ground truth data (i.e., the test data set) to compare against the forecasting results.

The maximum speed of the 5-day data for the freeway basic segment is 65.0385 mph, the minimum speed is 29.9108 mph, the average speed is 55.78036 mph, and the standard deviation is 4.92808.

Extraction of IMF Components

Following the EMD steps outlined in the section on Stage 1, a total of 10 IMFs and one residue were obtained from the original speed data series, as is shown in Figure 6. Figure 6 shows that the original speed series is decomposed into a series of relatively stationary speed data sets, which will be modeled 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 extracted in the last few components, such as IMF 7 and IMF 8. The highly stochastic variations in the speed series are represented by the first few components; the cyclical components in the speed series are shown by the last few components. The last component (depicted as residue) is the residue of the sifting process, which shows the overall trend of the speed series.

The original time series can be reconstructed by using the sifting process backward. By doing so, the intrinsic meaning of these IMF components can also be examined. The short-period components are illustrated in the first few IMFs, and the long-period components are illustrated 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.

Table 1 presents the descriptive statistics of the 10 intrinsic mode functions and the residue; the original traffic speed signal can be retrieved by the combination of each IMF component and residue. [See Hamad et al. on how to retrieve the original signal from the decomposed IMFs (*1*).]

Building the ARIMA Model

The ARIMA model has been widely used in time series predictions since it was initially proposed by Box and Jenkins (*3*). For example, it has been applied to predict traffic flow by Lee and Fambro (*13*) and Williams et al. (*16*). The results of both showed that the use of a subset of the ARIMA model would increase the accuracy of the short-term forecasting task in time series models. In addition, a hybrid short-term forecasting method called KARIMA was developed by Van Der Voort et al. to show the improved performance over single ARIMA models (*14*). All of these show that the ARIMA model has proved to be an effective method in short-term predictions.

To build an ARIMA model for each IMF, start with identifying the correlation (sample autocorrelation and partial autocorrelation functions) by using time series analysis. The autocorrelation function describes the correlation between values of a time series at different times as a function of the two time points or the time lag. Let $x(t)$ be the time series and i be some point in time after the start of the time series. Then x_s is the value at time s. Suppose the time series has mean μ_s, variance σ_s^2, and standard deviation σ_s, then the definition of the autocorrelation between time point s and t is

$$R(s,t) = \frac{E[(x_t - \mu_t)(x_s - \mu_s)]}{\sigma_t \sigma_s}$$

The partial autocorrelation function describes the correlation between series values that are k intervals apart. Accounting for the values of

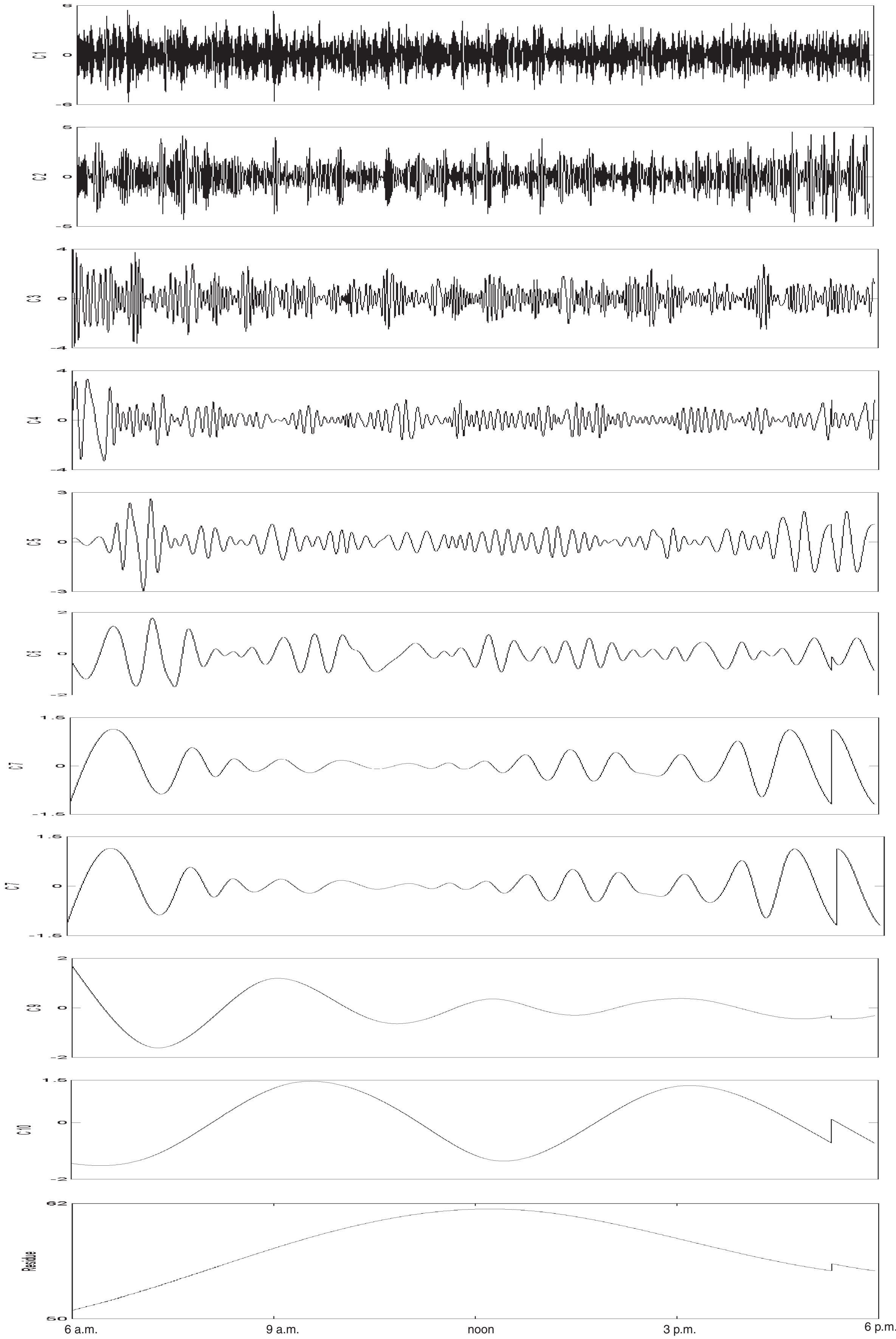

FIGURE 6 IMFs and residue extracted through EMD.

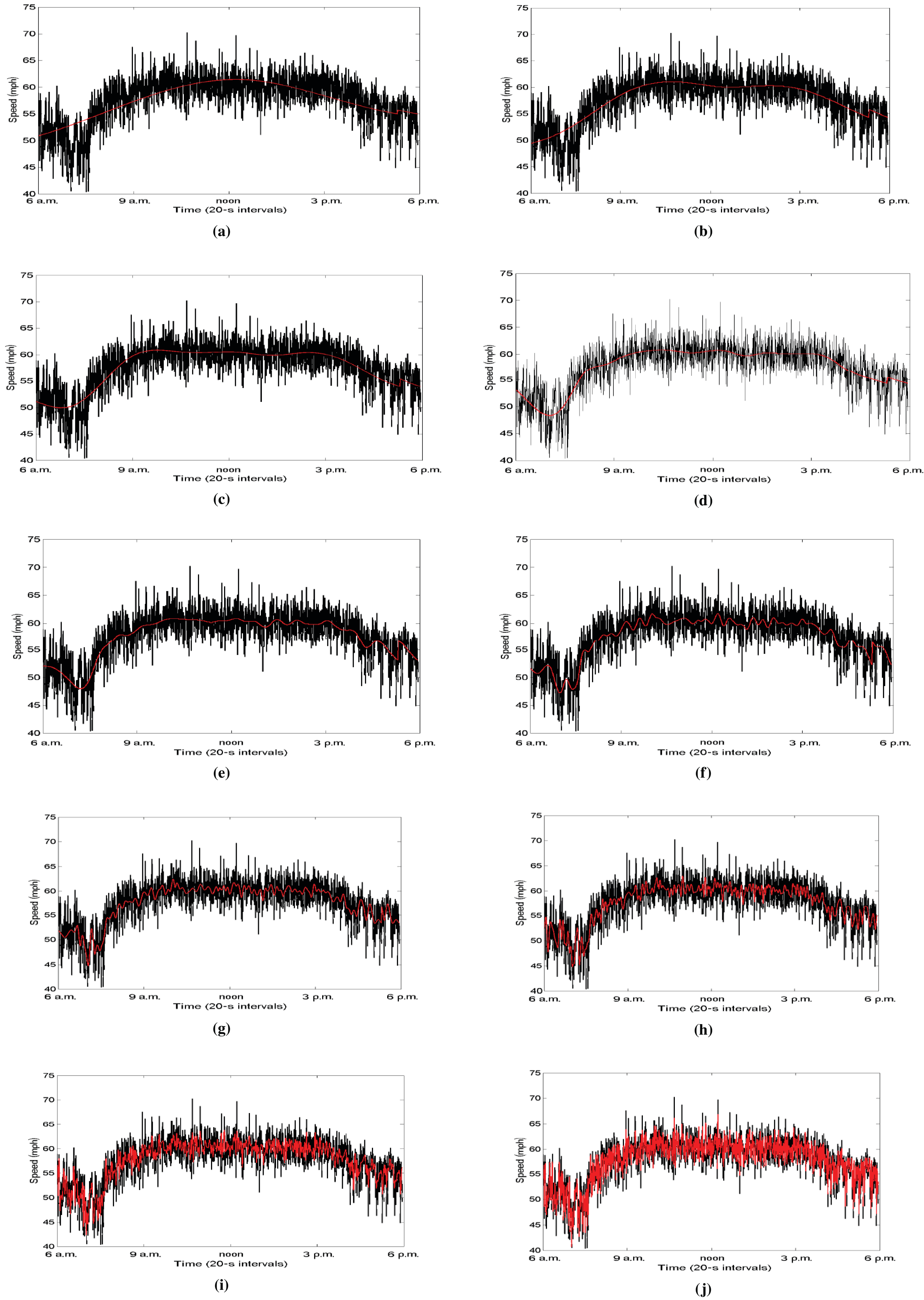

FIGURE 7 Reconstruction processes of original time series from IMFs: (*a*) original and C11, (*b*) original and sum of C11 to C10, (*c*) original and sum of C11 to C9, (*d*) original and sum of C11 to C8, (*e*) original and sum of C11 to C7, (*f*) original and sum of C11 to C6, (*g*) original and sum of C11 to C5, (*h*) original and sum of C11 to C4, (*i*) original and sum of C11 to C3, and (*j*) original and sum of C11 to C2.

TABLE 1 Descriptive Statistics of Each IMF and Residue

Component	Minimum	Maximum	Mean	Standard Deviation
Original	40.369	70.248	57.546	4.509
IMF1	–5.734	5.441	0.011	1.858
IMF2	–4.575	4.523	–0.007	1.429
IMF3	–3.890	3.945	0.015	1.012
IMF4	–3.322	3.328	0.011	0.786
IMF5	–2.980	2.620	0.002	0.683
IMF6	–1.610	1.740	–0.030	0.566
IMF7	–1.190	1.140	0.026	0.459
IMF8	–1.890	2.080	–0.061	0.635
IMF9	–0.081	1.720	–0.081	0.636
IMF10	–1.522	1.458	0.032	0.974
Residue	50.900	61.421	57.628	2.953

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. The partial correlation function α can be defined as follows:

$$\alpha(1) = \operatorname{cor}(z_t, z_{t+1})$$

$$\alpha(k) = \operatorname{cor}\left(z_{t+k} - P_{t,k}(z_{t+k}), z_t - P_{t,k}(z_t)\right) \qquad \text{for } k \geq 2$$

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

ANALYSIS OF RESULTS

In this section, a number of case studies are presented on short-term traffic speed forecasting with the proposed EMD-ARIMA modeling framework for freeway basic segments and freeway on-ramp and off-ramp. For each individual scenario, the results generated from the EMD-ARIMA framework will be compared against the Holt–Winters model, the traditional ARIMA model (ARIMA), and the ANN model to show its performance for different scenarios.

Performance Measures

The MAE, RMSE, and MAPE are calculated for each case to evaluate the accuracy of each model:

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|x_t - \hat{x}_t\right|$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(x_t - \hat{x}_t\right)^2}$$

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x_t - \hat{x}_t}{x_t}\right|$$

where

x_t = observation at the tth time interval,
$\hat{x}_t$ = predicted value at the tth time interval, and
N = number of observations.

The Holt–Winters model, the traditional ARIMA, and the ANN model are used as the benchmarking forecasting methods to justify and evaluate the performance of the proposed EMD-ARIMA model. Because of space constraints, details of the Holt–Winters model, the traditional ARIMA, and ANN methods are intentionally omitted.

Case Study: Basic Freeway Segment

The data were collected from a basic freeway segment on GA-400 through all of 2003; empirical traffic speeds were taken in the morning peak hours (7:00 to 9:00 a.m.) to conduct the analysis. The original data were recorded every 20 s. The forecast horizon was divided into four categories: 5, 10, 15, and 20 min; the forecasting results are provided in Figures 8 through 11.

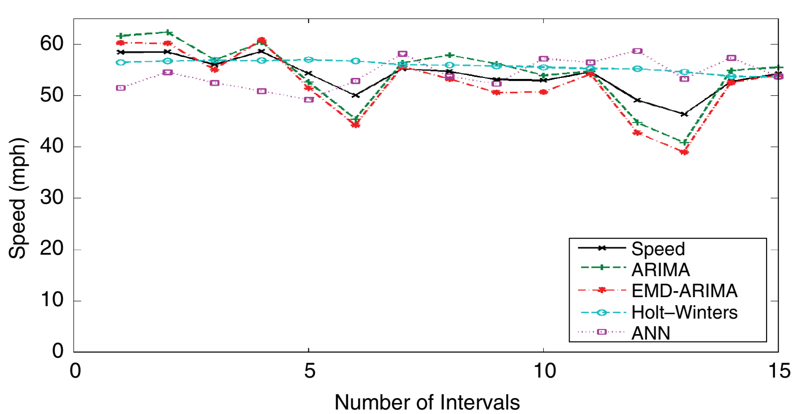

FIGURE 8 Forecasting results on freeway basic segments for 5-min horizons.

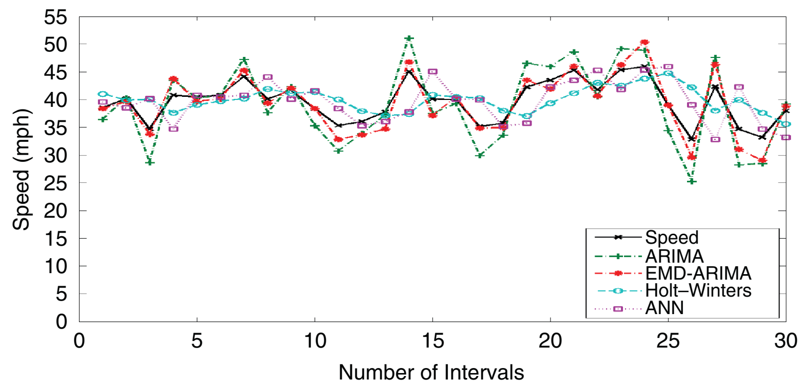

FIGURE 9 Forecasting results on freeway basic segments for 10-min horizons.

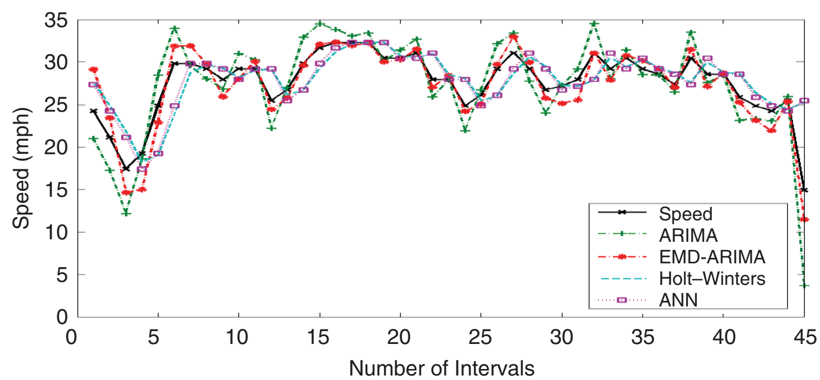

FIGURE 10 Forecasting results on freeway basic segments for 15-min horizons.

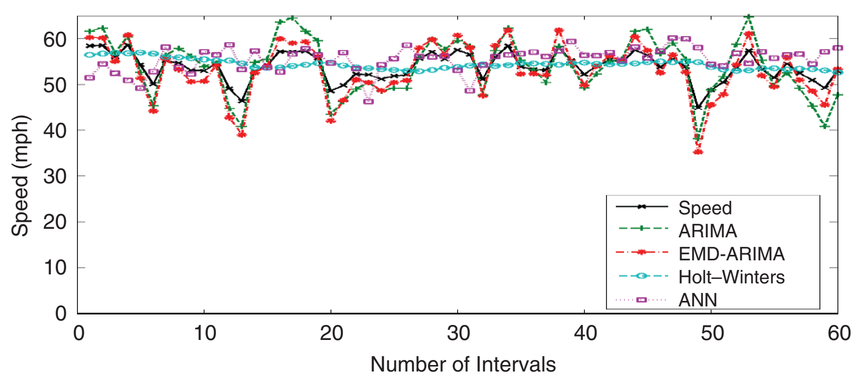

FIGURE 11 Forecasting results on freeway basic segments for 20-min horizons.

Table 2 presents the numerical comparison between the proposed EMD-ARIMA modeling framework and benchmarking forecasting methods (ARIMA, Holt–Winters, and ANN) with respect to MAE, MAPE, and RMSE. Results show that (*a*) the performance of the EMD-ARIMA model is superior to that of the ARIMA model, the Holt–Winters model, and the ANN model, especially for the extreme values; (*b*) the performance of the Holt–Winters model is second best, but it is close to the ARIMA model; (*c*) the EMD part of the hybrid model improves the performance of the traditional ARIMA model significantly (the MAPE values of the EMD-ARIMA model for the 5-, 10-, 15-, and 20-min forecasting horizons are 3.39%, 3.20%, 4.09%, and 3.84%, respectively, compared with the MAPE values of the ARIMA model, 6.78%, 5.68%, 5.38%, and 5.32%); and (*d*) compared with other models, the EMD-ARIMA model produces smaller prediction errors consistently, regardless of the single or multistep short-term forecasting.

CONCLUSIONS AND FUTURE WORK

A hybrid EMD-ARIMA model is developed and tested for short-term traffic speed prediction in the basic freeway segment scenario for four distinct time horizons, 5, 10, 15, and 20 min. Results show that (*a*) the proposed hybrid model generates more accurate prediction results than the traditional ARIMA model, the Holt–Winters model, and the ANN model, and (*b*) the performance is consistent in different

TABLE 2 Analysis of Model Performance for 5-, 10-, 15-, and 20-min Forecasting Horizons

Measure of Effectiveness	EMD-ARIMA	ARIMA	Holt–Winters	ANN
5-min Forecasting Results				
MAE (mph)	2.434	2.533	2.639	3.128
RMSE (mph)	2.958	3.289	3.510	3.848
MAPE (%)	4.742	4.822	5.149	5.833
10-min Forecasting Results				
MAE (mph)	2.363	2.975	2.670	3.128
RMSE (mph)	3.001	3.430	3.510	3.288
MAPE (%)	4.526	5.571	5.092	5.833
15-min Forecasting Results				
MAE (mph)	2.158	2.682	3.401	3.345
RMSE (mph)	2.747	3.183	2.971	4.067
MAPE (%)	4.071	4.982	4.517	6.270
20-min Forecasting Results				
MAE (mph)	2.218	2.956	2.412	3.299
RMSE (mph)	2.877	3.547	3.108	4.158
MAPE (%)	4.255	5.562	4.617	6.206

scenarios, and it can be applied to short-term freeway traffic speed predictions.

In the future, this hybrid model will be applied to other scenarios such as work zone traffic flow prediction and passenger flow prediction. In addition, it is worth investigating which model is more reliable in the hybrid model families (EMD-ARIMA, EMD-ANN, and EMD-SVM). The model will be tested with more data (or less data) to discuss its sensitivity to the amount of data. Another point worth exploring further is the model performance over different forecasting horizons. There are more unknowns than knowns concerning the hybrid forecasting models, that is, the computation time and the affordability of different methods.

REFERENCES

1. Hamad, K., M.T. Shourijeh, E. Lee, and A. Faghri. Near-Term Travel Speed Prediction Utilizing Hilbert–Huang Transform. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 24, 2009, pp. 551–576.
2. Chatfield, C. The Holt–Winters Forecasting Procedure. *Applied Statistics,* Vol. 27, No. 3, 1978, pp. 264–279.
3. Box, G.E.P., and G.M. Jenkins. *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco, Calif., 1976.
4. Bao, Y., T. Xiong, and Z. Hu. Forecasting Air Passenger Traffic by Support Vector Machines with Ensemble Empirical Mode Decomposition and Slope-Based Method. *Discrete Dynamics in Nature and Society,* Vol. 2012, 2012.
5. Liu, H., C. Chen, H. Tian, and Y. Li. A Hybrid Model for Wind Speed Prediction Using Empirical Mode Decomposition and Artificial Neural Networks. *Renewable Energy,* Vol. 48, 2012, pp. 545–556.
6. Okolobah, V., and Z. Ismail. Forecasting Peak Load Demand by ARIMA and EMD. *Archives des Sciences,* Vol. 66, No. 1, 2013.
7. Smith, B.L., and M.J. Demetsky. Short-Term Traffic Flow Prediction: Neural Network Approach. In *Transportation Research Record 1453,* TRB, National Research Council, Washington, D.C., 1994, pp. 98–104.
8. Dougherty, M.S., and H.C. Kirby. The Use of Neural Networks to Recognize and Predict Traffic Congestion. *Traffic Engineering Control,* Vol. 34, No. 6, 1998, pp. 311–314.
9. Castro-Neto, M., Y.-S. Jeong, M.-K. Jeong, and L.D. Han. Online-SVR for Short-Term Traffic Flow Prediction Under Typical and Atypical Traffic Conditions. *Expert Systems with Applications,* Vol. 36, No. 3, Part 2, 2009, pp. 6164–6173.
10. Ye, Z., and Y. Zhang. Short-Term Traffic Flow Forecasting Using Fuzzy Logic System Methods. *Journal of Intelligent Transportation Systems,* Vol. 12, No. 3, 2008, pp. 102–112.
11. Papageorgiou, M., and Y. Wang. Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter: A Case Study. *Transportation Science,* Vol. 42, No. 2, 2007, pp. 167–181.
12. Moorthy, C., and B. Ratcliffe. Short Term Traffic Forecasting Using Time Series Methods. *Transportation Planning and Technology,* Vol. 12, No. 1, 1988, pp. 45–56.
13. Lee, S., and D.B. Fambro. Application of Subset Autoregressive Integrated Moving Average Model for Short-Term Freeway Traffic Volume Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1678,* TRB, National Research Council, Washington, D.C., 1999, pp. 179–188.
14. Van Der Voort, M., M. Dougherty, and S. Watson. Combining Kohonen Maps with ARIMA Time Series Models to Forecast Traffic Flow. *Transportation Research Part C: Emerging Technologies,* Vol. 4, No. 12, 1996, pp. 307–318.
15. Kamarianakis, Y., and P. Prastacos. Forecasting Traffic Flow Conditions in an Urban Network: Comparison of Multivariate and Univariate Approaches. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1857,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 74–84.
16. Williams, B.M., P.K. Durvasula, and D.E. Brown. Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models. In *Transportation Research Record 1644,* TRB, National Research Council, Washington, D.C., 1998, pp. 132–141.
17. Wei, Y., and M.-C. Chen. Forecasting the Short-Term Metro Passenger Flow with Empirical Mode Decomposition and Neural Networks. *Transportation Research Part C: Emerging Technologies,* Vol. 21, No. 1, 2012, pp. 148–162.
18. Jiang, X., and H. Adeli. Dynamic Wavelet Neural Network Model for Traffic Flow Forecasting. *Journal of Transportation Engineering,* Vol. 131, No. 10, 2005, pp. 771–779.
19. Innamma, S. Short-Term Prediction of Travel Time Using Neural Networks on an Interurban Highway. *Transportation,* Vol. 32, No. 6, 2005, pp. 649–669.
20. Watson, S.M., H.R. Kirby, and M.S. Dougherty. Should We Use Neural Networks or Statistical Models for Short-Term Motorway Traffic Forecasting? *International Journal of Forecasting,* Vol. 13, No. 1, 1997, pp. 43–50.
21. Xie, Y., Y. Zhang, and Z. Ye. Short-Term Traffic Volume Forecasting Using Kalman Filter with Discrete Wavelet Decomposition. *Computer-Aided Civil and Infrastructure Engineering,* Vol. 22, No. 5, 2007, pp. 326–334.
22. Abdulhai, B., H. Porwal, and W. Recker. Short-Term Traffic Flow Prediction Using Neuron-Genetic Algorithm. *ITS Journal,* 2002, pp. 3–41.
23. Ngo, L.B., A. Apon, and D. Hoffman. An Empirical Study on Forecasting Using Decomposed Arrival Data of an Enterprise Computing System. Presented at 9th International Conference on Information Technology: New Generations, IEEE, Las Vegas, Nev., 2012, pp. 756–763.
24. Huang, N.E., et al. The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis. *Proc., Royal Society of London A,* Vol. 454, 1998, pp. 903–995.

The Artificial Intelligence and Advanced Computing Applications Committee peer-reviewed this paper.