Sensors and Materials, Vol. 36, No. 12 (2024) 5411–5425
MYU Tokyo

S & M 3876

Extreme-point Symmetric Mode Decomposition-based Sequential Data Assimilation System for Short-term Traffic Flow Prediction

Zhilin Wang,[1*] Zhanhai Zhang,[2] and Yuan Tian[3]

[1]School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, No. 1 Zhanlanguan Road, Beijing 100048, P.R. China
[2]Architecture Design and Research Institute of Beijing University of Civil Engineering and Architecture Co., Ltd., No. 1 Zhanlanguan Road, Beijing 100048, P.R. China
[3]Shanghai Topcon-Sokkia Technology & Trading Co., Ltd., No. 389, Gangao Road, Shanghai Pilot Free Trade Zone, Shanghai 200120, P.R. China

(Received March 28, 2024; accepted November 29, 2024)

Keywords: sequential data assimilation, ESMD, historical data denoising, short-term traffic flow prediction

Short-term traffic flow prediction plays an important role in intelligent transportation systems (ITSs). Sequential data assimilation (SDA) is very effective in the short-term traffic flow prediction of expressways because of its real-time reflections of local fluctuations of fast-changing traffic flow values in the time and space domains. Assimilation models in a traditional SDA (T-SDA) system are usually constructed using historical measurements. However, historical data are always disturbed by local noises, greatly affecting the accuracy of constructed assimilation models and predicted results. To deal with the problem, we propose to adopt the extreme-point symmetric mode decomposition (ESMD) method to conduct historical data denoising for improving the assimilation model performance in the SDA system. First, the original historical measurement signals are decomposed into a series of simple signals called intrinsic mode functions (IMFs) by ESMD to further analyze and seek useful information and local stochastic noises. Second, the denoised historical traffic data are used to construct an assimilation model, and the denoised SDA (D-SDA) system for short-term traffic flow prediction is established. Third, the applications of the D-SDA system for short-term traffic flow prediction are presented and compared with those of the T-SDA system. Experimental results showed that compared with the T-SDA system, the D-SDA system can successfully reduce the effects of noises in historical measurements on assimilation model construction and improve the accuracy of short-term traffic flow prediction results.

1. Introduction

With rapid economic development, the convenience of transportation has become a key factor restricting the development of urbanization. Intelligent transportation systems (ITSs) can alleviate traffic congestion by strengthening traffic flow guidance.[1,2] Short-term traffic flow

*Corresponding author: e-mail: wangzhilin@bucea.edu.cn
https://doi.org/10.18494/SAM5055

ISSN 0914-4935
https://myukk.org/

prediction can provide effective data support for traffic flow control and guidance in ITSs through real-time reflections of local fluctuations of fast-changing traffic flow values in the time and space domains.[3,4] Thus, how to acquire accurate short-term traffic flow prediction information is important to ensure effective traffic operation. It attracted much attention to take the advantages of various measurements and models to make the predictions.[5] Among them, sequential data assimilation (SDA) techniques based on Bayesian theory, as one of the implementation classes of data assimilation (DA), have been effectively used in short-term traffic predictions.[6,7] The statement parameters can be estimated *a posteriori* on the basis of status updating by referring to the weights of the model and measurement errors when measurements are available.[8] By considering the traffic state data distribution in both time and space, SDA can estimate the following traffic state vectors by integrating physical model information and measurements. In addition, the errors of historical and current measurements, together with the model background field errors, are calculated in each step to correct the short-term traffic prediction model. Three key components are necessary in the SDA system, namely, assimilation models (dynamic state model and observation model), measurements (historical and current), and assimilation methods.[9] For short-term traffic flow prediction, the expression of the SDA system can be shown mathematically as

$$\begin{cases} X_{state}^{i} = D^{i,i-1} X_{state}^{i-1} + C^{i,i-1} \varepsilon^{i-1}, \\ y_{meas}^{i} = M^{i} X_{state}^{i} + \delta^{i}, \end{cases} \tag{1}$$

where i denotes the discrete time index. Most studies show that data used in short-term traffic flow predictions are commonly aggregated from 1 to 30 min intervals.[10–12] In this study, traffic flow data are collected at 15 min intervals. The equation $X_{state}^{i} = D^{i,i-1} X_{state}^{i-1} + C^{i,i-1} \varepsilon^{i-1}$ describes the dynamic state model, which expresses the evolution of the traffic state parameters. $D^{i,i-1}$ is the dynamic state model recording the change of state parameter X_{state} from time index $i-1$ to i. $y_{meas}^{i} = M^{i} X_{state}^{i} + \delta^{i}$ is the dynamic observation model, which operates as the real-time connection between state parameter X_{state} and measurements y_{meas} at time index i through the time-dependent measurement operator M. ε and δ are assumed to be zero-mean Gaussian random noises with covariance matrices Q and R, respectively. C is a coefficient matrix.

As stated above, the SDA system for short-term traffic flow prediction mainly contains two steps, that is, statement parameter evolution based on the previous analyzed and statement parameter for next time interval updated under Bayes' formula.[13] Equation (1) is the conceptual assimilation model. Then, it should be considered how to build the dynamic state and observation models in detail. As variation patterns in historical measurements are similar on the same day of consecutive weeks or months, they are always used to construct forecast models such as the vector autoregressive (VAR) model.[9] The VAR model considers the effect of downstream and upstream location information on the specific location traffic flow. Also, because the Kalman filter (KF) method can update variable states using real-time measurements and adapt to changes in traffic flow, it becomes the basic algorithm in SDA systems. With the advantages of low computational cost and low storage requirements, KF has excellent performance in many

traffic flow prediction applications.[14,15] Thus, the VAR model will be referred to as the assimilation model and the KF method will be selected as the assimilation method in order to construct the traditional SDA (T-SDA) system for short-term traffic flow prediction. However, unavoidable random variation noises in historical traffic flow measurements always make it difficult to accurately extract the patterns of changes in traffic flow, which are indispensable for assimilation model construction, that is, the VAR model. Obviously, if the assimilation model is inaccurate, the final short-term traffic flow prediction accuracy will also be affected. Therefore, it is necessary to deal with the noises existing in historical traffic before the assimilation process.

The commonly used denoising methods can be divided into noise reduction in the time, frequency, and time–frequency domains. The denoising methods in the time domain, such as the KF and Chebyshev filter methods, are mainly based on mathematical operations to process discrete signal point data and then filter out the noise information.[16] However, these methods do not consider the statistical characteristics of the measured data during processing. They are mostly applicable to linear, stationary, and regular signals. Thus, they may be not suitable for processing nonlinear, nonstationary, and irregular time series historical traffic flow data. The denoising methods in the frequency domain convert the sampled signal to the frequency domain for analysis and processing. From the spectrum difference between the effective signal and the noise signal, the truncated frequency threshold can be obtained, then the noise spectrum can be removed, and noise reduction can be achieved. Fourier transform series methods are more commonly used because of their advantages of simple operation, flexible frequency selection, effective avoidance of time shift, and good noise reduction effect.[17,18] However, the selection of the threshold of truncation frequency to distinguish effective signals and noise is mostly based on manual experience in Fourier transform method applications. Moreover, the elimination of noise information of different frequency scales is difficult using only one single truncation frequency threshold. The denoising methods in the time–frequency domain can process and analyze nonlinear nonstationary signals in the two-dimensional time–frequency domain. Among them, empirical mode decomposition (EMD) series algorithms are commonly used, such as the EMD method,[19] ensemble EMD (EEMD) method,[20] and extreme-point symmetric mode decomposition (ESMD) method.[21] The EMD method can adaptively decompose a complex signal into a series of intrinsic mode functions (IMFs) with different frequencies, which contain different time-scale characteristics of the source signal. The method can break through the time-domain limitations of the frequency domain noise reduction method and realize random noise reduction by eliminating the low-order IMF component with high frequency. However, the mode aliasing and end effects arise. Also, the trend function of EMD is relatively rough, which severely limits the effect of signal noise reduction. After studying the statistical characteristics of EMD and white Gaussian noise signals, the EEMD method is proposed; it can alleviate the influence of the mode aliasing effect on the signal noise reduction accuracy. However, the EEMD method is susceptible to the influence of additional noise amplitude and the number of integrated experiments. It is also difficult to reduce the influence of instantaneous noise. Considering the limitations of the EMD and EEMD methods, in the ESMD method, the internal extreme symmetric difference is used to form a signal envelope. In addition, the least squares method is used to optimize the final residual mode to obtain the highest adaptive global mean

(AGM), which can effectively reduce the difficulty of determining the mode decomposition screening times and the effect of mode aliasing. In the noise reduction of nonstationary and nonlinear monitoring signals, the ESMD method can reduce the noise effect by eliminating a certain number of high-frequency and low-order IMFs.

Therefore, to solve the poor reliability problem of the assimilation model in T-SDA systems caused by unavoidable random variation noises in historical traffic flow measurements, an innovative denoised SDA (D-SDA) system is constructed and applied to short-term traffic flow prediction. The following three critical issues are investigated: (i) the original historical measurement signals are decomposed into a series of simple signals, IMFs, by ESMD to further analyze and seek the useful information and local stochastic noises, (ii) the denoised historical traffic data acquired are used to construct an assimilation model, and (iii) the D-SDA system for short-term traffic flow prediction is established, combining the denoised assimilation model, measurements, and standard KF method. The remainder of the paper is organized as follows. The T-SDA system for short-term traffic flow prediction is introduced in Sect. 2. The denoising processing of historical measurement signals and the construction of the D-SDA system are described in Sect. 3. In Sect. 4, the applications of the D-SDA system for short-term traffic flow prediction are presented and compared with those of the T-SDA system. Finally, conclusions are given in Sect. 5.

2. T-SDA System for Short-term Traffic Flow Prediction

As described earlier, there are three necessary components of the T-SDA system: assimilation models (dynamic state model and observation model), measurements (historical and current), and assimilation methods. The VAR model is the assimilation model and the KF method is selected as the assimilation method. The measurements are of traffic flow values. The standard KF method is used as the assimilation method in the T-SDA system. The VAR model can be expressed as

$$\frac{flow_s^{i+1}}{flow_s^{i+1}} = pa_0^i \times F^i + pa_1^i \times F^{i-1} + pa_2^i \times F^{i-2} + \ldots + pa_n^i \times F^{i-n} + \delta(i) \tag{2}$$

with

$$F^i = \begin{bmatrix} \dfrac{flow_s^i}{flow_s^i} & \dfrac{flow_{A_j}^i}{flow_{A_j}^i} \end{bmatrix}^T \left(j = 1,2,3,\ldots,m\right), \tag{3}$$

where $[pa_0^i, pa_1^i, \ldots, pa_n^i]$ are unknown state parameters in the dynamic state model. $flow_s^{i+1}$ is the traffic flow value on a specific path and needs to be predicted using the T-SDA system. $\overline{flow_s^i}$ denotes the average of a specific path and can be acquired from historical traffic flow measurements in the time interval $[iT, (i + 1)T]$. T represents the sample interval of 15 min. As

variation patterns in historical measurements are similar on the same day of consecutive weeks or months, they are always used to construct forecast models. The historical measurements mentioned here are traffic flow data from the same day in previous weeks. To be specific, if predicted traffic flow values on Monday need to be acquired and analyzed, traffic flow data sets of the former seven consecutive Mondays are selected and used for model construction in the T-SDA system. $flow_s^i$ is the historical traffic flow value of the specific path in the time interval $[(i-1)T, iT]$ and $\overline{flow_s^i}$ is the corresponding average. $flow_{A_j}^i$ denotes the traffic flow value of the downstream and upstream paths, which are also the paths adjacent to the predicted path in the time interval $[(i-1)T, iT]$. j is the number of adjacent paths. $\overline{flow_{A_j}^i}$ is the corresponding historical average.

Then, the T-SDA system for short-term traffic flow prediction can be built by combining Eq. (1) with Eqs. (2) and (3). Information in Eq. (1) can be expressed in detail by setting

$$\begin{cases} D^{i,i-1}=\mathbf{I}, C^{i,i-1}=\mathbf{0}, \\ X_{state}^i=\left[pa_0^i, pa_1^i, ..., pa_n^i\right]^T, \\ y_{meas}^i=\dfrac{flow_s^{i+1}}{\overline{flow_s^{i+1}}}, \\ M^i=\left[F^{iT}, F^{(i-1)T}, F^{(i-2)T}, ..., F^{(i-n)T}\right]. \end{cases} \tag{4}$$

As mentioned above, the assimilation method is the key component in the T-SDA system, which connects assimilation models and measurements. The standard KF method is used as the assimilation method in this study as it is effective under both stationary and nonstationary conditions. The KF method is a well-known technique to track state values over time. Its efficient calculations and small storage requirements make it more appropriate for short-term traffic flow forecasting.[7,14] The standard KF method used as the assimilation method in the T-SDA system can be expressed as the forecast part shown by Eq. (5) and the update part shown by Eq. (6).

$$\begin{cases} X_{state\_}^i f=D^{i,i-1}X_{state\_}^{i-1}a \\ P_{state\_}^i f=D^{i,i-1}P_{state\_}^{i-1}a\left(D^{i,i-1}\right)^T \end{cases} \tag{5}$$

$$\begin{cases} K^i=P_{state\_}^i f\left(M^i\right)^T\left(M^iP_{state\_}^i f\left(M^i\right)^T+R^i\right)^{-1} \\ X_{state\_}^i a=X_{state\_}^i f+K^i\left(y_{meas}^i-M^iX_{state\_}^i f\right) \\ P_{state\_}^i a=\left(I-K^iM^i\right)P_{state\_}^i f \end{cases} \tag{6}$$

Here, $P^i_{state\_}f$ denotes the error covariance matrix of the state vector prediction values, and $P^i_{state\_}a$ is the error covariance matrix of the estimated state vector values. As stated above, R^i denotes the error covariance matrix of the Gaussian random noise series of the observation equation, as shown in Eq. (1). K^i is the Kalman gain matrix, which is crucial for balancing the weight between state estimates and new measurements.

3. D-SDA System Construction

It can be seen from Eq. (4) that the measurement operator M^i is acquired from historical measurements. The measurement operator M^i will also play an important role in Kalman gain matrix calculation in the KF method, as shown in Eq. (6). Therefore, to improve the accuracy of measurement operator construction and assimilation calculation, a denoising process for historical measurements is essential before short-term traffic flow prediction using the SDA system. The ESMD method can decompose the complex historical traffic signal into a collection of band-limited IMFs by determining an optimal global mean curve in an adaptive way.[22,23] Different IMFs represent different physical meanings with inherent natural frequencies from high to low frequency.[23] For a given historical traffic signal $flow_h(i)$, the main steps of ESMD decomposition are shown as follows.

(1) Obtain all local minima and maxima value points of historical traffic signal $flow_h(i)$, represented as $extre_c$ with c = 1, 2, ..., ω. List the midpoints of the adjacent extreme points as mid_c with c = 1, 2, ..., $\omega - 1$. Considering that extreme points cannot exist at both ends of historical traffic signal $flow_h(i)$, a waveform feature matching method is used to extend the historical traffic signal $flow_h(i)$ and suppress the end effect.

(2) Apply the second-order odd–even curve interpolation and third-order B-spline curve interpolation to fit the midpoints of adjacent extreme points. The midpoint curve interpolation model is determined using the residual standard deviation of fitted data. The calculation model of the weighted midpoint curve $\bar{L}$ is established using the number of midpoints involved in the calculation.

(3) Introduce the optimal sifting times term Ψ_0 and permitted error ε to guarantee the optimal IMF decomposition and ensure the quality of decomposition (too few modal decomposition times may reduce the symmetry of the decomposed IMF, while too many modal decomposition times may destroy the inherent amplitude change of the signal and result in a decomposed IMF without physical meaning). The historical traffic signal $flow_h(i)$ can be decomposed into q IMFs and one adaptive average curve r as

$$flow_h^l(i) = \sum_{k=1}^{q} IMF_k^l(i) + r^l(i) \quad l = 1, 2, \cdots g, \tag{7}$$

where l means the lth historical measurement signal of measurement group g, which is needed to be decomposed.

(4) Identify the IMF dominated by high-frequency noise and the IMF dominated by useful information by the Spearman correlation coefficient method. Then, remove the IMF component dominated by high-frequency noise. The signal after high-frequency noise reduction can be obtained by reconstructing the remaining IMF components and adaptive average curve *r*. Thus, the denoised historical traffic signal $Dflow_h(i)$ can be acquired as

$$Dflow_h^l(i) = \sum_{k=k_{th}+1}^{q} IMF_k^l(i) + r^l(i) \quad l = 1,2,\cdots g, \tag{8}$$

where k_{th} is the turning point between the IMF dominated by high-frequency noise and the IMF dominated by useful information.

(5) The instantaneous noise and most of the high-frequency noise in the historical traffic measurement values can be eliminated. Replace $flow_s^i$ with $Dflow_s^i$ in Eqs. (2)–(4). Then, the DA model in the D-SDA system can be constructed as

$$\begin{cases} D^{i,i-1} = \mathbf{I}, C^{i,i-1} = \mathbf{0}, \\ X_{state}^i = \left[pa_0^i, pa_1^i, ..., pa_n^i \right]^T, \\ y_{meas}^i = \dfrac{Dflow_s^{i+1}}{\overline{Dflow_s^{i+1}}}, \\ M^i = \left[F^{iT}, F^{(i-1)T}, F^{(i-2)T}, ..., F^{(i-n)T} \right], \end{cases} \tag{9}$$

with

$$F^i = \begin{bmatrix} \dfrac{Dflow_s^i}{\overline{Dflow_s^i}} & \dfrac{Dflow_{A_j}^i}{\overline{Dflow_{A_j}^i}} \end{bmatrix}^T \left(j = 1,2,3,...,m \right). \tag{10}$$

As stated above, the D-SDA system for short-term traffic flow prediction can be established, and the entire technological schematic framework is shown in Fig. 1.

4. Experiments and Discussion

4.1 Effectiveness analysis of noise reduction using ESMD

As the ESMD method is selected to process the random variation noises in historical traffic flow measurements because of its good performance, the simulation experiment is firstly conducted to verify the effectiveness of ESMD denoising in this study. The simulation signal $s(i)$ consists of useful signals $s_1(i)$ and $s_2(i)$ and a random noisy part $\kappa(i)$ with an SNR of 20 dB built

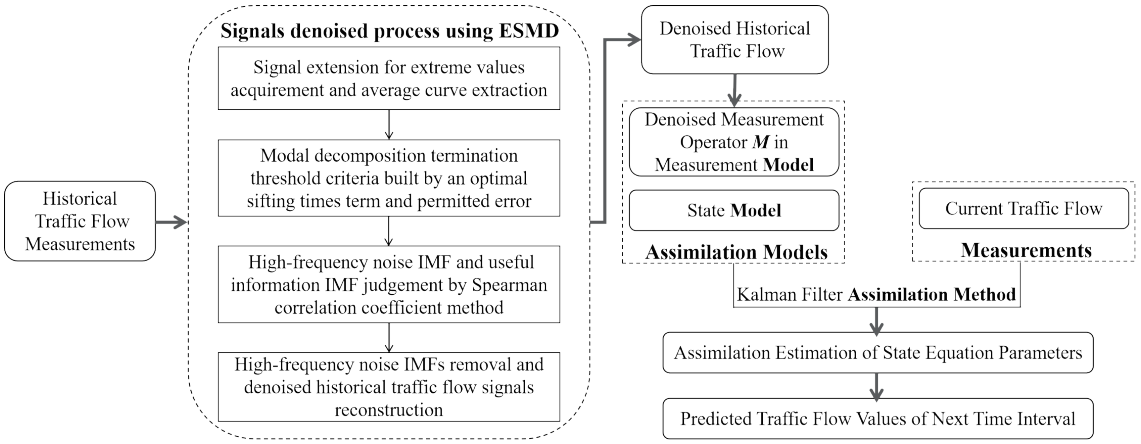

Fig. 1. (Color online) Schematic of D-SDA system for short-term traffic flow prediction.

to reflect the change characteristics of historical traffic flow measurement. The data sampling frequency is 200 Hz and the number of sampling points is 600. The specific information of each signal is given as Eq. (11) and the corresponding waveforms are shown in Fig. 2.

$$\begin{cases} s(i) = s_1(i) + s_2(i) + \kappa(i) \\ s_1(i) = 0.6 \times \sin(6\pi i) \\ s_2(i) = \cos(15\pi i + 2\sin(2i)) \end{cases} \tag{11}$$

To prove the noise reduction accuracy of ESMD, the EMD and EEMD methods are selected to process the simulated data above. Also, the root mean square error (*RMSE*) and mean absolute percent error (*MAPE*)[9,21] are employed to evaluate the denoising performance of each method. Figure 3 shows the denoised results acquired using the EMD, EEMD, and ESMD methods. Table 1 lists *RMSE* and *MAPE* values of the denoised results acquired using the three methods. It can be seen from Fig. 3 that the denoised line obtained by ESMD is much closer to the original one without noise. The EMD denoised line still has some unexpected fluctuations, which may be a result of mode aliasing or the end node effect. Also, a different good performance is produced in terms of *RMSE* and *MAPE* values. For example, *RMSE* and *MAPE* are 0.0538 and 0.2574 for ESMD noise reduction, respectively. They are reduced by 22.92 and 32.93%, respectively, compared with those for EMD. They are also smaller than those of EEMD. The simulated results show that the ESMD method is not only feasible and effective in random noise reduction, but also has a higher denoising effect than either the EMD or EEMD denoising method.

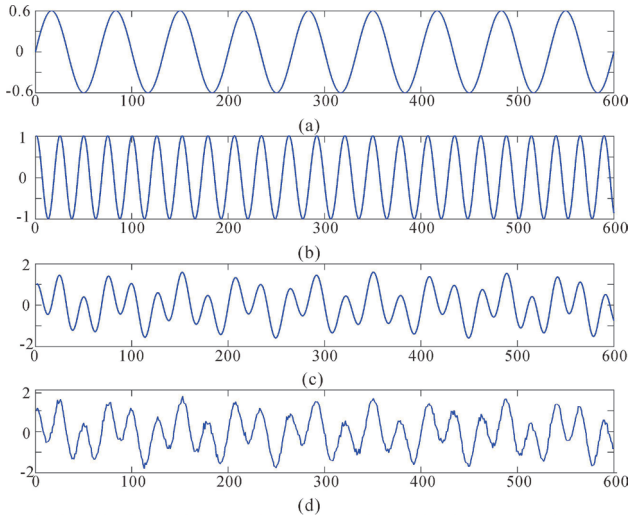

Fig. 2. (Color online) (a), (b) Signal components $s_1(i)$ and $s_2(i)$, respectively. (c), (d) Simulation signals with and without noise, respectively.

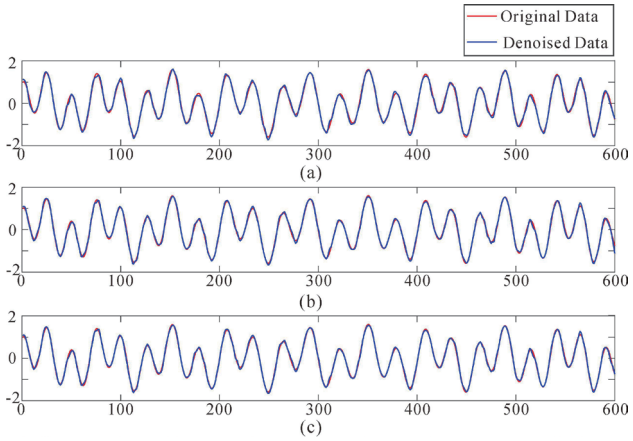

Fig. 3. (Color online) Noise reduction results acquired by (a) EMD, (b) EEMD, and (c) ESMD methods.

Table 1
Noise reduction results of three methods.

	EMD	EEMD	ESMD
RMSE	0.0698	0.0590	0.0538
MAPE	0.3838	0.3151	0.2574

*The smaller the *RMSE* and *MAPE* index values, the higher the noise reduction effect.

4.2 Effectiveness analysis of D-SDA system on short-term traffic prediction

4.2.1 Practical area and material description

To further detect the effectiveness of the proposed D-SDA system for short-term traffic flow prediction, a practical experiment is conducted. The study area is a subarea of the highway between Liverpool and Manchester, UK, as shown in Fig. 4(a). The traffic flow measurements

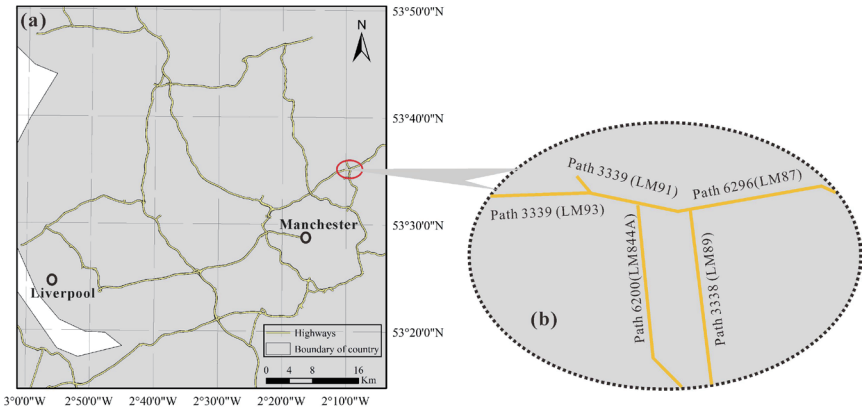

Fig. 4. (Color online) (a) Study area and (b) part of the paths.

are downloadable from the website of Highways England (*highwaysengland.co.uk*). The time interval for the datasets is 15 min. Several adjacent paths, as shown in Fig. 4(b), are selected as representatives to verify the validity of the proposed method. They are path 3339 (LM91), path 6200 (LM844A), path 3339 (LM93), path 3338 (LM89), and path 6296 (LM87). Traffic flow prediction results of each path are acquired and analyzed from Monday to Sunday. The traffic flow data of each path contains eight days from consecutive weeks. Datasets of the first seven days are used for the construction of assimilation models in the SDA system, and the datasets from the eighth day are employed to test the effectiveness of the proposed D-SDA system.

4.2.2 Impacts of denoised historical traffic data on prediction results

We take the short-term traffic flow prediction for path 3339 (LM91) as a detailed example to illustrate the impacts of denoised historical traffic data on the improvement of the accuracies of the assimilation model and forecasting results. There are four paths: path 3338 (LM89), path 3339 (LM93), path 6200 (LM844A), and path 6296 (LM87). All are adjacent to path 3339 (LM91), as shown in Fig. 4(b). Obviously, $m = 4$ in Eq. (10). n is set to be 2 in Eq. (9). First, the original traffic flow data of path 3339 (LM91) on three consecutive Mondays, which have similar variations but are affected by observation noises, are shown in Fig. 5(a). These original data series are decomposed into purer and noisy series using the ESMD method, as shown in Figs. 5(b) and 5(c), respectively. Similarly, original traffic flow data of path 3339(LM91) on three consecutive Saturdays with purer and noisy series obtained by the ESMD method are shown in Figs. 5(d)–5(f). It can be seen from Fig. 5 that images of the purer series are smoother than those of the original series but still retain their variation trends. They can be treated as denoised data for further multiscale model building. Then, to verify the impacts of denoised historical traffic data on prediction results, traffic flow prediction results of path 3339 (LM91) obtained on the workday Monday and the non-workday Saturday are analyzed in detail without loss of generality. For the comparative analysis of experimental results, prediction results are obtained from the T-SDA system using the raw data without the noise reduction. Besides, as the traffic flows early

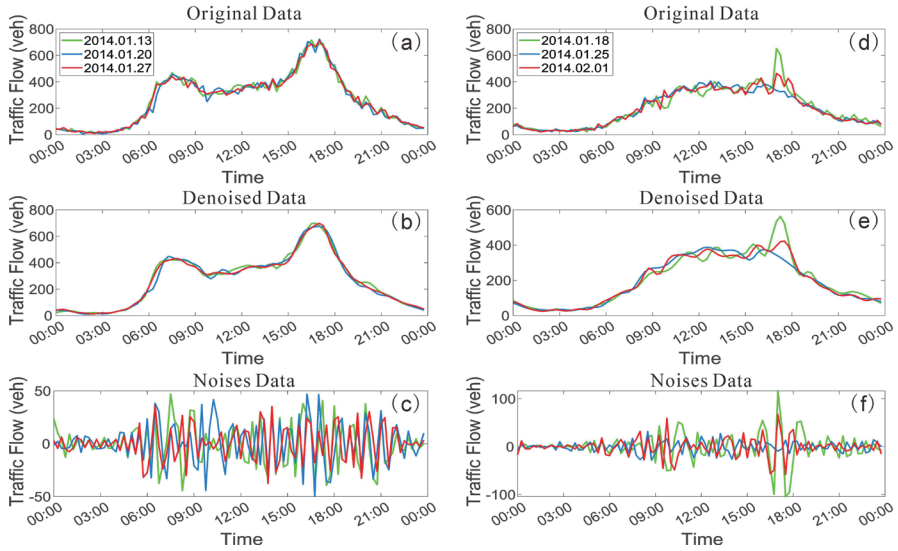

Fig. 5. (Color online) (a)–(c) Original traffic data, purer series, and noise on three consecutive Mondays. (d)–(f) Original traffic data, purer series, and noise on three consecutive Saturdays.

in the morning and late at night are small and of little concern to traffic management, only traffic flow prediction results from 6:00 a.m. to 9:00 p.m. are acquired and analyzed. Figures 6(a) and 6(b) show the traffic flow prediction results of path 3339 (LM91) on the workday Monday and on the non-workday Saturday, respectively. The prediction results from the T-SDA system are taken as a reference to verify the effectiveness of assimilation prediction results obtained from the D-SDA system. Meanwhile, the true traffic flow values are also added. Table 2 shows the corresponding *RMSE* and *MAPE* values of the short-term traffic flow prediction results from the T-SDA and D-SDA systems on Monday and Saturday, respectively.

It can be seen from the results shown in Fig. 6 and Table 2 that the prediction results acquired from the D-SDA system using historical measurements with noise reduction are superior to those obtained from the T-SDA system. The lower *RMSE* and *MAPE* values of D-SDA are shown in Table 2, and the distributions displayed in Fig. 6 are better than those from T-SDA. For example, the *RMSE* and *MAPE* of the prediction results are respectively 37.43 and 7.88% in the case of the D-SDA system on Monday. They were reduced by 6.57% (from 44.00 to 37.43) and 1.61% (from 9.49 to 7.88%) compared with the results from T-SDA. Similar prediction results were obtained using the data collected on Saturday. It can be drawn that random noises in historical traffic flow measurements indeed affect the precision of assimilation model construction and assimilation prediction results. The D-SDA system proposed in this study is effective in improving the short-term traffic flow prediction accuracy.

4.2.3 D-SDA system performance for short-term traffic flow prediction

To further verify the effectiveness of the proposed D-SDA system, the short-term traffic flow prediction of all the paths shown in Fig. 4(a) is conducted from Monday to Sunday. As an example of a detailed analysis, the prediction performance characteristics of the five paths shown in Fig. 4(b) are listed first. The *RMSE* and *MAPE* values of the prediction results acquired

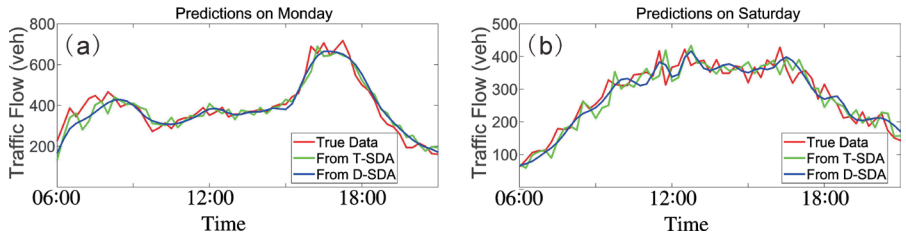

Fig. 6. (Color online) Different traffic flow prediction performance characteristics of path 3339 (LM91) from the (a) T-SDA system and (b) D-SDA system on Monday and Saturday.

Table 2
RMSE and *MAPE* of prediction results from T-SDA and D-SDA on Monday and Saturday.

	Monday		Saturday	
	RMSE	*MAPE* (%)	*RMSE*	*MAPE* (%)
T-SDA system	44.00	9.49	35.89	10.10
D-SDA system	37.43*	7.88*	26.86*	8.74*

*Highest performance

from the T-SDA and proposed D-SDA systems from Monday to Sunday are given in Fig. 7, and the average *RMSE* and *MAPE* values of the five paths are shown in Tables 3 and 4, respectively.

As shown in Tables 3 and 4, compared with the prediction results obtained from the T-SDA system, the average *RMSE* and *MAPE* were improved by various degrees when using the proposed D-SDA system. Taking path 3339 (LM93) as an example, the average *RMSE* and *MAPE* acquired from the T-SDA system were 29.96 and 10.53%, respectively. A higher prediction performance is acquired from the D-SDA system where the average *RMSE* was reduced by 7.41 (from 29.96 to 22.55) and the relative accuracy was improved by 24.73%. The corresponding average *MAPE* was reduced by 2.43% (from 10.53 to 8.10%) and the relative accuracy was improved by 23.08%. Similar results were also obtained for the other four paths. The results of the test indicate that the assimilation models built using historical measurements with noise reduction can achieve a higher prediction accuracy.

For further verification, the T-SDA and proposed D-SDA systems are applied for the short-term traffic flow prediction of all the paths shown in Fig. 4(a). The corresponding average *RMSE* and *MAPE* values from Monday to Sunday are given in Tables 5 and 6, respectively. Results showed that the average *RMSE* and *MAPE* values from the D-SDA system are all smaller than those from the T-SDA system. Taking the workday Monday and non-workday Sunday as examples, the average *RMSE* and *MAPE* from the T-SDA system on Monday are 79.69 and 9.50%, respectively. The corresponding values are 70.06 and 8.01% from the D-SDA system, respectively. On the non-workday Sunday, the *RMSE* and *MAPE* values from the T-SDA system were 47.25 and 9.76%, and the corresponding values were 36.97 and 7.82% from the D-SDA system, respectively.

Overall, the results in Fig. 7 and Tables 3–6 suggest that the performance of the D-SDA system is higher than that of the T-SDA system on improving the accuracy of assimilation model construction and short-term traffic flow predictions. The proposed D-SDA system is proved to be effective and valid in short-term traffic flow prediction.

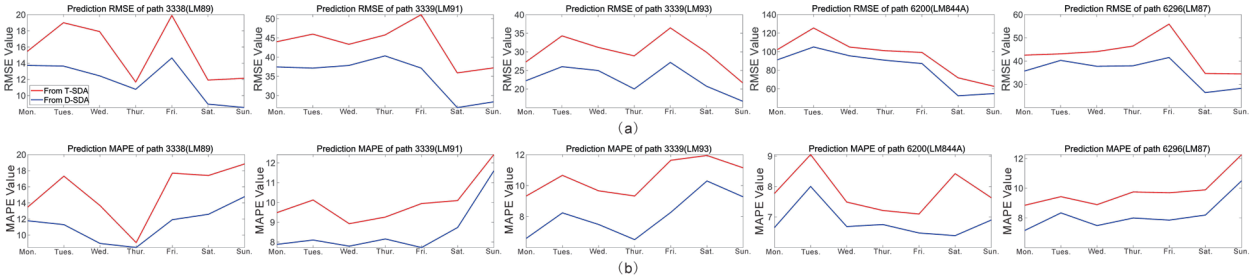

Fig. 7. (Color online) (a) *RMSE* and (b) *MAPE* values of five paths from Monday to Sunday.

Table 3
Average *RMSE* values of five paths under T-SDA and D-SDA systems.

Path	3338 (LM89)	3339 (LM91)	3339 (LM93)	6200 (LM844A)	6296 (LM87)
T-SDA	15.43	43.32	29.96	95.36	43.03
D-SDA	11.81*	35.01*	22.55*	82.48*	35.41*

*Highest performance

Table 4
Average *MAPE* values of five paths for five models.

Path	3338 (LM89)	3339 (LM91)	3339 (LM93)	6200 (LM844A)	6296 (LM87)
T-SDA	15.35	10.04	10.53	7.81	9.82
D-SDA	11.40*	8.56*	8.10*	6.84*	8.22*

*Highest performance

Table 5
Average *RMSE* values of all paths for five models.

	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.	Sun.
T-SDA	79.69	83.04	83.89	83.12	84.47	50.84	47.25
D-SDA	70.06*	73.56*	73.92*	72.35*	72.89*	39.74*	36.97*

*Highest performance

Table 6
Average *MAPE* values of all paths for five models.

	Mon.	Tues.	Wed.	Thur.	Fri.	Sat.	Sun.
T-SDA	9.50	9.63	10.17	9.35	9.40	9.32	9.76
D-SDA	8.01*	8.24*	8.62*	7.83*	7.87*	7.63*	7.82*

*Highest performance

5. Conclusions

In this study, aiming to reduce the effect of random noise in historical traffic flow measurements on assimilation model construction and prediction results, we adopted the ESMD method to conduct the denoising process to improve the accuracy of assimilation model construction and built a D-SDA system for short-term traffic flow prediction. The results demonstrated that the proposed D-SDA system has the ability to perform well in short-term

traffic flow prediction applications. Specifically, the results presented in this paper clearly highlighted the following:

(1) Noise can be successfully separated and dealt with from historical measurements using the ESMD method. Compared with denoised results from EMD or EEMD, the noise reduction effect of ESMD is the highest.

(2) The built D-SDA system can be successfully applied to short-term traffic flow prediction. The prediction results acquired from the D-SDA system outperformed those from the T-SDA system. For the traffic flow prediction of path 3339 (LM93), the accuracy can be increased by 24.73%. For all paths, the prediction accuracy can be increased by up to 21.83%.

In this study, the time interval was set as 15 min for highway short-term traffic flow prediction. In future work, the proposed method can be tested and applied to traffic flow prediction in urban areas with more complex traffic conditions. Thus, the optimization and improvement of the current D-SDA system are necessary for shorter-term traffic flow prediction.

Acknowledgments

We thank Highways England for providing all the traffic flow open data.

References

1 A. Barodi, A. Zemmouri, A. Bajit, M. Benbrahim, and A. Tamtaoui: Microprocess. Microsyst. **100** (2023) 104830. https://doi.org/10.1016/j.micpro.2023.104830
2 X. Luo, D. Li, and S. Zhang: J. Sens. **2019** (2019) 1. https://doi.org/10.1155/2019/6461450
3 S. A. Zargari, S. Z. Siabil, A. H. Alavi, and A. H Gandomi: Expert Syst. **29** (2012) 124. https://doi.org/10.1111/j.1468-0394.2010.00567.x
4 Q. Ren, Y. Li, and Y. Liu: Expert Syst. Appl. **227** (2023) 120203. https://doi.org/10.1016/j.eswa.2023.120203
5 G. W. Inverarity, W. J. Tennant, and L. Anton: Q. J. R. Meteorolog. Soc. **149** (2023) 1138. https://doi.org/10.1002/qj.4431
6 M. A. Makridis and A. Kouvelas: Transp. Res. Part C Emerging Technol. **149** (2023) 104066. https://doi.org/10.1016/j.trc.2023.104066
7 J. Guo, W. Huang, and B. M. Williams: Transp. Res. Part C Emerging Technol. **43** (2014) 50. https://doi.org/10.1016/j.trc.2014.02.006
8 S. Symon, D. Sipp, P. J. Schmid, and B. J. Mckeon: AIAA J. **58** (2020). https://doi.org/10.2514/1.j057889
9 R. Wang, W. Shi, X. Liu and, Z. Li: ISPRS Int. J. Geo-Inf. **9** (2020) 731. https://doi.org/10.3390/ijgi9120731
10 Y. Xu, H. Chen, Q. J. Kong, X. Zhai, and Y. Liu: J. Adv. Transport. **50** (2016) 489. https://doi.org/10.1002/atr.1356
11 S. Dunne and B. Ghosh: J. Transp. Eng. **138** (2012) 455. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000337
12 M. Voort, M. Dougherty, and S. Watson: Transport. Res. C-Emer. **4** (2012) 307. https://doi.org/10.1016/s0968-090x(97)82903-8
13 A. Fournier, G. Hulot, D. Jault, W. Kuang, A. Tangborn, N. Gilet, E. Canet, J. Aubert, and F. Lhuillier: Space. Sci. Rev. **155** (2010) 247. https://doi.org/10.1007/s11214-010-9669-4
14 Y. Xie, Y. Zhang, and Z. Ye: Comput-Aided. Civ. Inf. **22** (2007) 326. https://doi.org/10.1111/j.1467-8667.2007.00489.x
15 M. Deng and S. Qu: Comput. Intel. Neurosc. **2015** (2015). https://doi.org/10.1155/2015/875243
16 J. Yu, X. Meng, X. Shao, B. Yan, and L. Yang: Eng. Struct. **81** (2014) 432. https://doi.org/10.1016/j.engstruct.2014.10.010
17 W. Shi and R. Wang: Sens. Mater. **32** (2020) 3893. https://doi.org/10.18494/SAM.2020.2969
18 M. Farooq Wahab, Fabrice Gritti, and Thomas C. O'Haver: TrAC. Trends Anal. Chem. **143** (2021) 116354. https://doi.org/10.1016/j.trac.2021.116354
19 G. Han, B. Lin, and Z. Xu: J. Instrum. **12** (2017) P03010. https://doi.org/10.1088/1748-0221/12/03/P03010
20 J. Li, Y. Tong, L. Guan, S. Wu, and D. Li: RSC Adv. **8** (2018) 8558. https://doi.org/10.1039/C7RA13202F

21 R. Wang, Y. Huang, X. Liu, H. Wang, M. Jiang: Sens. Mater. **34** (2022) 4001. https://doi.org/10.18494/SAM4008
22 Y. Yang, F. Xiong, Z. Wang, and H. Xu: Int. J. Struct. Stab. Dyn. **20** (2020) 718. https://doi.org/10.1142/S0219455420500455
23 Y. Li and C. Yue: Water Sci. Technol. Water Supply. **20** (2020) 1439. https://doi.org/10.2166/ws.2020.058