Sensing Data Supported Traffic Flow Prediction via Denoising Schemes and ANN: A Comparison

Xinqiang Chen, Shubo Wu, Chaojian Shi, Yanguo Huang, Yongsheng Yang, Ruimin Ke, Jiansen Zhao

***Abstract*—Short-term traffic flow prediction plays a key role of Intelligent Transportation System (ITS), which supports traffic planning, traffic management and control, roadway safety evaluation, energy consumption estimation, etc. The widely deployed traffic sensors provide us numerous and continuous traffic flow data, which may contain outlier samples due to expected sensor failures. The primary objective of the study was to evaluate the use of various smoothing models for cleaning anomaly in traffic flow data, which were further processed to predict short term traffic flow evolution with artificial neural network. The wavelet filter, moving average model, and Butterworth filter were carefully tested to smooth the collected loop detector data. Then, the artificial neural network was introduced to predict traffic flow at different time spans, which were quantitatively analyzed with commonly-used evaluation metrics. The findings of the study provide us efficient and accurate denoising approaches for short term traffic flow prediction.**

***Index Terms*—traffic flow prediction, artificial neural network, data denoising, Butterworth filter.**

I. Introduction

The rapid development of urbanization has intensified gap between huge traffic demand and traffic facility shortage, which results in serious traffic problems (e.g., traffic congestion, traffic accident). Traffic flow prediction provides crucial information for microscopic and macroscopic traffic state estimation, which attracts increasing interest in the transportation research community. Currently, various traffic flow prediction models have been proposed which show numerous successes (i.e., high accuracy). The traffic prediction models comprise of mathematic based and machine learning supported frameworks. More specifically, traditional mathematic models mainly employ statistical and calculus relevant methods to fulfill the traffic flow prediction task. The commonly used models involve exponential smooth [1, 2], Kalman filter [3-6], autoregressive integrated moving average (ARIMA) [7-10], etc. Note that traffic flow prediction accuracy is heavily relied on the raw traffic flow data quality. Thus, statistical based traffic flow prediction frameworks may contain unexpected anomalies without our data quality control procedure.

The machine learning based methods can correct out potential traffic flow data noises due to the intrinsic advantage of reasoning capability. Previous studies employ support vector machine (SVM) [11-13], back propagation (BP) neural networks [14, 15] and fuzzy neural networks [16, 17] to extract embedded traffic flow patterns, and thus the noisy data samples are corrected in an automatic manner. More specifically, the anomaly outliers will be replaced with smoothed samples via nonlinear functions. Tang et al., proposed an improved Markov Chain framework to estimate traffic flow data series considering spatial-temporal correlation links [18]. Lv et al., proposed a stacked autoencoder model to learn distinct features form traffic flow data, which is the trial of employing deep learning model to tackle the traffic flow prediction task [19]. Hosseini introduced a convolutional neural network based traffic flow model to prediction traffic flow volume with a time-space diagram [20]. The machine learning based traffic flow prediction has become a hot topic due to the advantage of high accuracy, less human-being involvement, etc. [21-27]

Previous studies suggested that prediction accuracy may be deteriorated due to noises existed in the raw traffic flow data, and many studies have been conducted to obtain the noise-free data [28, 29]. Peng et al., proposed a hybrid framework to predict short term traffic flow data via artificial neural network and genetic algorithm [30]. The experimental results indicated that the prediction performance of the proposed model is better than the ensemble empirical model decomposition based prediction models. Sun et al., developed a two-layer Fourier transform based traffic flow prediction framework, which was further tested on England Highway data [31]. Habtemichael et al., proposed a non- parametric model for the purpose of short-term traffic forecasting, which was implemented by identifying traffic patterns from big transportation data with an improved K-nearest neighbor (KNN) algorithm [32]. Xia et al., proposed a spatial-temporal weighted KNN model with a general purpose MapReduce framework to fulfill the accurate and efficient short-term traffic flow prediction [33]. Zhang et

Manuscript was received on March 4, 2020. This work was jointly supported by the National Natural Science Foundation of China (51709167, 51708094, 51579143), Shanghai Committee of Science and Technology, China (18040501700, 18295801100, 17595810300).

Xinqiang Chen and Yongsheng Yang are with the Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai, 201306, China (email: chenxinqiang@stu.shmtu.edu.cn, yangys@shmtu.edu.cn)

Shubo Wu, Chaojian Shi and Jiansen Zhao are with the Merchant Marine College, Shanghai Maritime University, Shanghai, 201306, China (email: wushubo@stu.shmtu.edu.cn, cjshi@shmtu.edu.cn, jszhao@shmtu.edu.cn)

Yanguo Huang is with School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou 341000, China; (jxhuangyg@126.com)

Ruimin Ke (corresponding author) is with the Department of Civil and Environmental Engineering, University of Washington, WA 98195, USA (email: ker27@uw.edu)

al., proposed a deep autoencoder neural network to forecast traffic congestion by extracting traffic parameters with computer vision techniques [34]. Similar studies can be found in [35-42]. In sum, data denoising procedure can significantly benefit traffic flow prediction accuracy.

Previous studies suggested that short term traffic flow prediction can be implemented in the time spans of 1 min, 2 min and 10 min [43, 44]. Following the rule, we implemented the traffic flow prediction task at same interval. Our primary contributions were summarized as follows: (1) we denoised the raw traffic flow data at different time intervals (i.e., 1 min, 2 min and 10 min) with various smoothing algorithms. More specifically, the data denoising procedure was implemented with wavelet filter (with different bases), moving average model (at different window sizes) and the Butterworth filter; (2) we forecasted the traffic flow variation tendency (at different time resolutions) with the artificial neural network (ANN); (3) we quantified the traffic flow prediction performance on the empirical traffic flow data (collected from three neighboring loop detectors on weekdays). The remainder of the paper is organized as follows. Section II introduces different denoising methods and ANN model in detail. Section III describes the data source, evaluation metrics and traffic flow data denoising procedure and prediction accuracy. Section IV briefly concludes the study and future work.

II. Methodology

The traffic flow data collected from on-spot sensors may be contaminated by various outliers, and data pre-processing is essential before implementing traffic flow analysis task. We employ three popular data-denoising models for the purpose of obtaining noise-free traffic flow data, which are illustrated in detail in the following sections.

A. Wavelet Filter

Wavelet filter (WL) decomposes traffic flow data into several subsets, which are termed as scaling and wavelet subsets. The wavelet relevant subsets contain details and noises in the original traffic flow data, and the latter are supposed to be suppressed during the data denoising procedure [45, 46]. Note that both of the wavelet basis and level of decomposition are very important for obtaining satisfied smoothing results (i.e., appropriate scaling and wavelet subsets). Our previous study suggested that wavelet basis plays a more important role in suppressing the traffic flow-like data [27-29]. Following the rule, we implement traffic flow denoising task with the WL filter performance under varied wavelet bases, and more details can be found in the experimental section. The original traffic flow data series can be formulated as Eq. (1). The WL model implements the traffic flow data denoising procedure via the following three steps:

Step 1: data decomposition; The raw traffic flow data is decomposed into several data series with the help of wavelet basis at a pre-defined decomposition level N. Note that the WL decomposition level may impose negative / positive effect on the model denoising performance.

Step 2: set appropriate thresholds for varied detail coefficients; The rule includes the hard and soft thresholding manners, which are shown in Eq. (2) and (3), respectively. We suppress the noisy data subsets while keeping the detail and noise-free subset by thresholding out the outliers.

Step 3: data reconstruction. We obtain the denoised traffic flow data by accumulating the wavelet coefficients subsets based on the previous step output. The efficient wavelet bases (including daubechies (db), coiflets (coif), symlets (sym) and haar) will be implemented in our study.

$$x_i = y_i + \varepsilon_i, i = 1,2,\cdots,n \tag{1}$$

$$\hat{\beta} = \begin{cases} \beta, & |\beta| \geq T \\ 0, & |\beta| < T \end{cases} \tag{2}$$

$$\hat{\beta} = \begin{cases} \mathrm{sgn}(\beta)(|\beta| - T), & |\beta| \geq T \\ 0, & |\beta| < T \end{cases} \tag{3}$$

where x_i is the empirical data sample obtained from the spot loop sensor, y_i is the noise-free data sample and ε_i is the noise for the current data sample x_i The parameter n is the volume for the traffic flow data series. The parameter β is wavelet coefficient, and $\hat{\beta}$ is estimated wavelet coefficient. The symbol is signum function and T is threshold.

B. Moving Average

Moving average (MA) is a traditional statistical method for analyzing time series data variation tendency, which can be used for data prediction and smoothing [47]. The MA model removes the anomaly data in the manner of averaging neighboring samples, which involves two types of smoothing logics. More specifically, the first type is symmetric moving average, which generates the target data by accumulating the forward and backward data samples. Note that each neighboring data sample shares same weight during averaging procedure. The second type is named as an asymmetric moving average, with each sample is assigned with different weights. The symmetric and asymmetric moving average models are calculated as follows:

$$\begin{cases} y_i = x_i, \ i \leq \frac{m-1}{2} \text{ or } i \geq n - \frac{m-1}{2} \\ y_i = \frac{1}{m}\left(\alpha_{i-\frac{m-1}{2}} x_{i-\frac{m-1}{2}} + \cdots + \alpha_{i+\frac{m-1}{2}} x_{i+\frac{m-1}{2}}\right), \ \frac{m-1}{2} < i < n - \frac{m-1}{2} \end{cases} \tag{4}$$

$$\begin{cases} y_i = x_i, \ i \leq m \\ y_i = \frac{1}{m}(\alpha_{i-m+1} x_{i-m+1} + \alpha_{i-m+1} x_{i-m+1} + \cdots + \alpha_i x_i), \ i > m \end{cases} \tag{5}$$

where parameter n is the total number of traffic flow data sample, α_i is the weight for the ith sample. The parameter m is the moving window size for the MA method, which should be an odd number in symmetric MA model. The MA smoothing performance is heavily relied on the window size. More specifically, larger window size may obtain over-smoothed traffic flow data (i.e., the data details may be wrongly suppressed), while smaller window size may fail to completely remove outliers in the traffic flow data series. In our study, we employ the simple symmetric moving average method (i.e., $\alpha_i = 1$) to denoise the initial traffic flow data.

The window size plays a crucial role for the MA model smoothing performance, and the parameter is set to typical values (i.e., 3, 5, 7, 9) to obtain holistic data denoising results.

Note that the symbol MA3 represents the moving average model with window size setting to 3. Following the rule, we employ the symbol MA5, MA7, MA9 to demonstrate window size setting to 5, 7 and 9 for the moving average model.

C. Butterworth Filter

Butterworth filter (abbreviated as BW) is first designed to suppress signal noises in the manner of obtaining maximal flat frequency response. Given a sampling frequency for the traffic flow data, the BW model accomplishes data denoising procedure by adjusting fluctuation magnitude. More specifically, the raw data samples with obvious abnormal oscillations will be replaced with smoothed data. The transformation function of BW model is shown in Eq. (6).

$$H(z) = \frac{a_0+a_1z^{-1}+\cdots+a_kz^{-k}}{b_0+b_1z^{-1}+\cdots+b_kz^{-k}} \quad (6)$$

where $a_i, b_i (i = 1,2,\cdots,k)$ are the coefficient set which determine the optimal response of BW model, and k presents the filter order. For the purpose of obtaining optimal traffic flow data denoising performance, we need to carefully determine the parameters cutoff frequency f_c and lowpass filter order kth. Note that larger f_c will produce flatter traffic flow data series, which may remove data details, and vice versa.

D. Artificial Neural Network

By obtaining noise-free traffic flow data, the Artificial neural network (ANN) is then introduced to predict traffic flow variation tendency in near future. The ANN involves static and dynamic neural network, which is identified via the memory storage mode. More specifically, the neural network is considered as a static neural network when the network does not contain memory (e.g., BP neural network). Moreover, the network involving memory storage is reckoned as dynamic neural network, such as Hopfield neural network, nonlinear autoregressive with external input (NARX) neural network. We employ the NARX model to predict traffic flow at different time span. The NARX network predicts traffic flow data with the input of data at previous timestamp and feedback. The formula for the NARX neural network is represented as follows [48-50]:

$$\hat{y}(t+1) = f\left(y(t),\cdots,y(t-k_y),x(t),\cdots,x(t-k_x)\right) \quad (7)$$

where parameters $x(t)$ is input of previous data sample the NARX neural network at time t, and $y(t)$ is feedback factor at corresponding time stamp. The k_x and k_y are the input and output memory orders, respectively, and $f(\cdot)$ is nonlinear function. The symbol $\hat{y}(t+1)$ is output for the NARX neural network at time $t+1$. The NARX neural network can obtain real-time traffic flow prediction performance due to the simplified model structure, which mainly contains input layer, hidden layer and the output layer (see Fig. 1). The NARX input layer contains both of the training and output traffic flow samples, which are denoted as x(t) and y(t), respectively. The hidden layer learns intrinsic traffic flow patterns from the input traffic flow samples. The parameters w and b (shown in the hidden and output layers in the NARX network) are the weight and threshold, respectively. In addition, the parameters s and r represent the delay orders for the x(t) and y(t), respectively.

E. Evaluation metrics

It is noted that the Root mean square error (RMSE) is widely used to measure the bias between the denoised and raw data, and the mean absolute error (MAE) can avoid disadvantage of data-deviation cancelation. More specifically, both of the Root mean square error (RMSE) and mean absolute error (MAE) can perfectly measure the bias between the denoised and ground-truth traffic data. In that manner, the statistics of RMSE and MAE are employed to quantify the traffic flow prediction accuracy. In addition, the mean absolute percentage error (MAPE) and signal-noise ratio (SNR) are used to evaluate the traffic flow data noise removal performance. The RMSE, MAE, MAPE and SNR are calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (8)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \quad (9)$$

$$MPAE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \quad (10)$$

$$SNR = 10\log_{10}\sum_{i=1}^{n}\frac{y_i^2}{(\hat{y}_i - y_i)^2} \quad (11)$$

where n is the volume of the traffic flow sample, and y_i is the ith sample of the raw data series, and $\hat{y}_i$ is counterpart of the ground truth data.

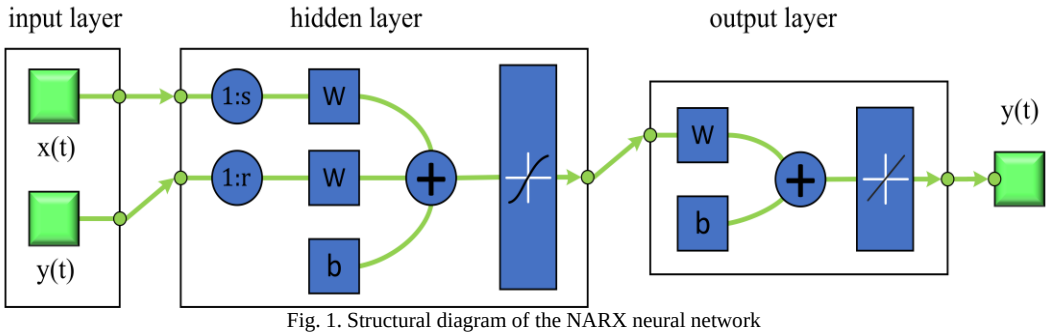


Fig. 1. Structural diagram of the NARX neural network

III. Experiments

We aim to verify traffic flow prediction performance via conventional ANN model (NARX neural network is used in our study) and hybrid ANN framework (i.e., combining various data denoising models before implementing the ANN prediction model). The delay orders in the ANN are both set to 2, and the number of hidden layer node is set to 5. The traffic flow prediction experiments were implemented on Windows 10 with Intel Core i7-8750H CPU @ 2.20GHz processor and RAM is 8G. The GPU version is NVIDA GeForce GTX 1050 Ti.

A. *Datasets*

We sampled the raw empirical traffic flow data from three loop deductive sensors (the sensor IDs are #5802, #5805 and #5808), which were deployed on the Interstate 494 in the Minnesota State, United States. The traffic flow data were collected from three consecutive weekdays from January 6, 2016 to January 8, 2016 for the purpose of avoiding holiday traffic variation interference. Moreover, we aggregated the raw traffic flow sensing data into different time interval (i.e., 1 min, 2 min and 10 min) [28]. More specifically, we collected 4320 samples at 1 min interval, while 2160 and 432 samples for 2 and 10 min interval, respectively. The traffic flow samples are considered as a two-dimension data series, which include time stamp and traffic flow volume.

B. *Data Denoising Performance Analysis*

The observed traffic flow data quality is crucial for traffic flow prediction accuracy, and thus data quality control is essential to smooth the noisy traffic flow data. To comprehensively compare varied denoising framework performance, we employ the WL model with different wavelet bases, MA and BW model to pre-process the raw data. The RMSE, MAE and SNR statistics help us analyze varied smoothing methods in a quantitative manner. Overall, there is no significant difference between varied smoothing models on the same time span data samples (e.g., smoothing performance on 1 min (2 min and 10 min) traffic flow data is quite similar among varied denoising models). Taken traffic flow denoising results on the 1 min data from sensor ID # 5802 as an example (see table 1), the WL (db) model obtains optimal noise removal performance in comparison with other wavelet bases denoising results (i.e., optimal RMSE, MAE and SNR for the wavelet models are 3.1197, 2.3615 and 14.4957).

We have tested the BW filter performance under varied parameter settings, which showed similar results. In that manner, we set the default cut-off frequency to 0.2. The BW model showed slightly-worse smoothing performance compared to those of the counterparts considering that the RMSE, MAE and SNR indicators are 3.6334, 2.7035 and 13.1714. The main reason is that the BW model considered the significant data oscillations as outliers, and thus some traffic flow details (caused by special events, etc.) may be wrongly suppressed. The MA models obtain better smoothing results in comparison with the WL and BW models. More specifically, the minimal RMSE and MAE for the MA model are 2.7257 and 2.0466, respectively, and maximal SNR is 15.6683, which are obtained when window size is set to 3. The smoothing results for 1 min data at sensor ID #5805 and #5808 (see table 2 and 3) show similar performance as those of sensor ID #5802.

Table 1. Statistical indices of different denoising methods for traffic data with sensor ID #5802

	1 min			2 min			10 min		
	RMSE	MAE	SNR	RMSE	MAE	SNR	RMSE	MAE	SNR
WL (db)	3.1197	2.3615	14.4957	4.3462	3.3166	17.5480	11.5172	8.8698	22.9919
WL (coif)	3.1477	2.3713	14.4178	4.4297	3.3803	17.3828	12.3668	9.4308	22.3737
WL (sym)	3.1487	2.3738	14.4152	4.3714	3.3349	17.4979	12.1911	9.7086	22.4980
WL (haar)	3.3023	2.4997	14.0014	4.2192	3.2759	17.8057	13.7382	10.9795	21.4603
MA3	2.7257	2.0466	15.6683	3.7924	2.8335	18.7320	8.9116	6.6775	25.2197
MA5	2.9653	2.2090	14.9363	4.0690	3.0519	18.1206	11.0291	7.9880	23.3681
MA7	3.0457	2.2662	14.7040	4.2332	3.1401	17.7770	12.2359	8.7249	22.4662
MA9	3.0837	2.2963	14.5964	4.3374	3.2226	17.5658	13.2333	9.3920	21.7855
BW	3.6334	2.7035	13.1714	5.5929	4.1623	15.3576	13.5891	10.0383	21.5551

Table 2. Statistical indices of different denoising methods for traffic data with sensor ID #5805

	1 min			2 min			10 min		
	RMSE	MAE	SNR	RMSE	MAE	SNR	RMSE	MAE	SNR
WL (db)	3.2787	2.4999	14.0359	4.5856	3.5057	17.0482	12.1269	9.3004	22.5019
WL (coif)	3.2487	2.4821	14.1160	4.6324	3.5202	16.9600	12.9141	10.0177	21.9556
WL (sym)	3.2504	2.4705	14.1114	4.5609	3.5036	17.0952	12.5328	9.8389	22.2159
WL (haar)	3.4114	2.6023	13.6914	4.9285	3.8039	16.4219	15.5079	12.6247	20.3658
MA3	2.8223	2.1122	15.3379	3.9123	2.8997	18.4275	11.3272	7.7840	23.0944
MA5	3.0786	2.2920	14.5831	4.2551	3.1713	17.6980	13.2669	9.1231	21.7215
MA7	3.1509	2.3448	14.3813	4.4631	3.2928	17.2834	14.0232	9.7688	21.2399
MA9	3.1895	2.3738	14.2757	4.6494	3.4238	16.9282	15.0454	10.5445	20.6288
BW	3.8839	2.8524	12.5646	6.1754	4.4655	14.4629	16.1489	11.1579	20.0140

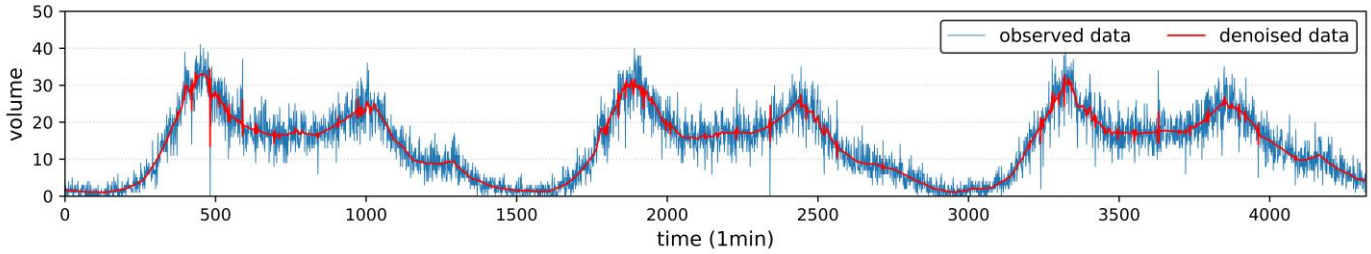
Table 3. Statistical indices of different denoising methods for traffic data with sensor ID #5808

	1 min			2 min			10 min		
	RMSE	MAE	SNR	RMSE	MAE	SNR	RMSE	MAE	SNR
WL (db)	3.0526	2.3477	13.9810	4.3230	3.3294	16.8835	11.4264	8.8416	22.3391
WL (coif)	3.0974	2.3720	13.8543	4.4901	3.4540	16.5543	11.1273	8.6987	22.5695
WL (sym)	3.0701	2.3528	13.9312	4.4160	3.4167	16.6988	11.4420	8.9656	22.3272
WL (haar)	3.2411	2.4699	13.4603	4.7457	3.6396	16.0734	12.4805	10.0452	21.5727
MA3	2.6234	1.9931	15.2968	3.7599	2.8343	18.0957	8.5570	6.6219	24.8509
MA5	2.8822	2.1736	14.4797	4.0716	3.0727	17.4041	10.4236	7.7759	23.1369
MA7	2.9764	2.2367	14.2004	4.1769	3.1588	17.1823	11.3132	8.3624	22.4256
MA9	3.0161	2.2738	14.0854	4.2715	3.2302	16.9876	12.2373	8.9259	21.7436
BW	3.5133	2.6556	12.7599	5.4060	4.0582	14.9418	12.7094	9.4452	21.4148

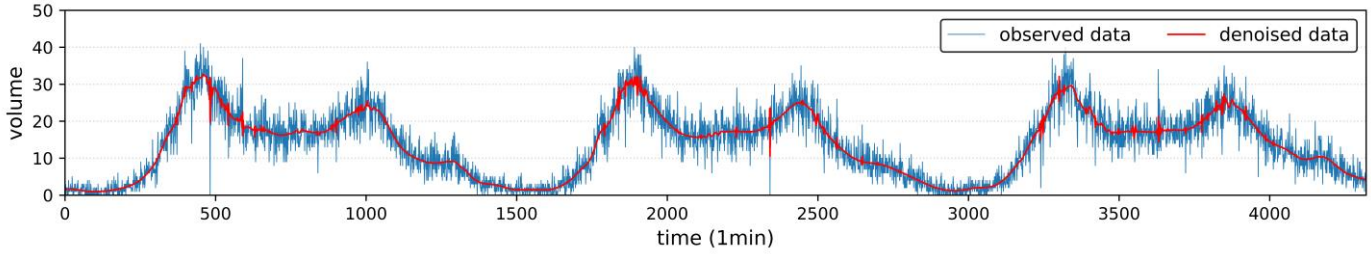
The smoothing performance on the 2 min and 10 min data series show a decreasing tendency by contrast with the 1 min traffic flow data. For instance, the minimal RMSE for the 2 min data (see table 1) is 3.7924 (obtained by MA3 model), minimal MAE is 2.8335 and maximal SNR is 18.7320. Though the maximal SNR is larger than that of 1 min data, the RMSE and MAE for the 2 min traffic flow data are both larger than the counterparts of 1 min. The 10-min RMSE and MAE values are three-fold higher than those of the 1 min, which indicates that traffic flow details in larger time span can be easily suppressed by the smoothing models.

It is noted that the WL(sym) model obtained better smoothing performance compared to the wavelet counterparts, which can be found in the traffic flow denoising performance at different time intervals (e.g., 1 min, 2 min and 10 min prediction results shown in table 2 and 3, respectively). Besides, the MA3 model obtained the optimal smoothing performance in the moving average families. More specifically, the average RMSE, MAE and SNR for sensor ID # 5805 (i.e., aggregating the 1 min, 2 min and 10 min data shown in table 2) are 6.0206, 4.2653 and 18.9533, and the counterparts for the sensor ID # 5808 are 4.9801, 3.8164 and 19.4145 (i.e., aggregating the 1 min, 2 min and 10 min denoising results in table 3). The traffic flow denoising performance obtained by the Butterworth model is significantly worse than the counterparts of the WA and the MA models (see the last row of the table 2 and 3, respectively).

To further examine denoising effects by the various models, we looked smoothing details of how each of the model address the outliers in the original traffic flow volume data. It is observed that the WL, MA and BW can successfully smooth the anomaly oscillations without discarding data details. Taken the denoising effect on data samples with sensor ID #5802 at 1 min scale as an example, the variation tendency was successfully shown in the denoised traffic flow data in each subplot in Fig.2. After carefully checking the denoising details of WL and BW, we found that several normal data oscillations were suppressed. Different window sizes used in the MA models (from 3 to 9) were tested, and the smoothing results suggested that the best performance was obtained when the window size was set to 3. The Fig. 3 and 4 showed similar smoothing result for the traffic flow data at sensor ID #5802 under time scales 2 min and 10 min. In sum, various smoothing models showed similar result on suppressing the data outliers, and the MA3 model obtained slightly better performance in comparison with other smoothing methods.

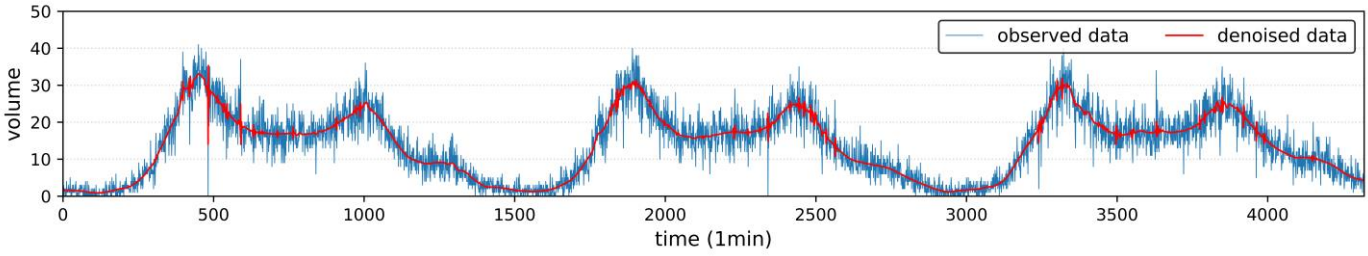


(a) denoising effect of WL (db)

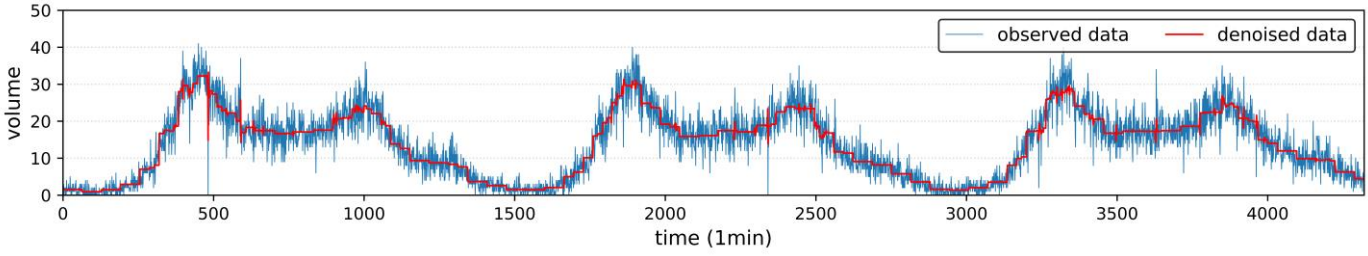


(b) denoising effect of WL (coif)

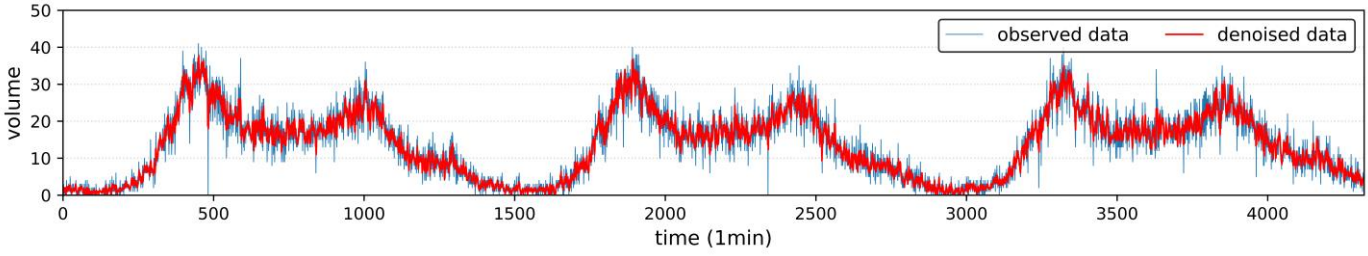
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2020.3007809, IEEE Sensors Journal



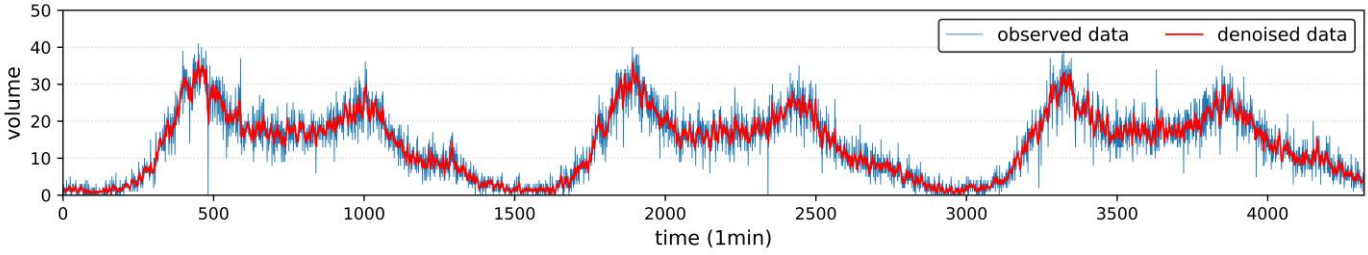
(c) denoising effect of WL (sym)



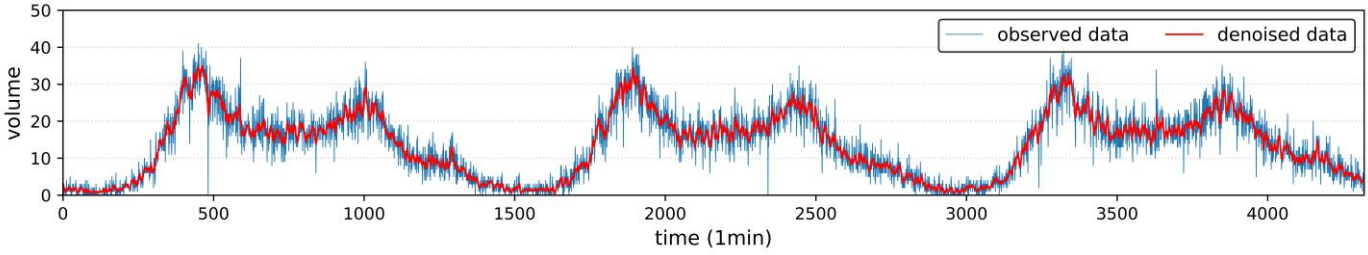
(d) denoising effect of WL (haar)



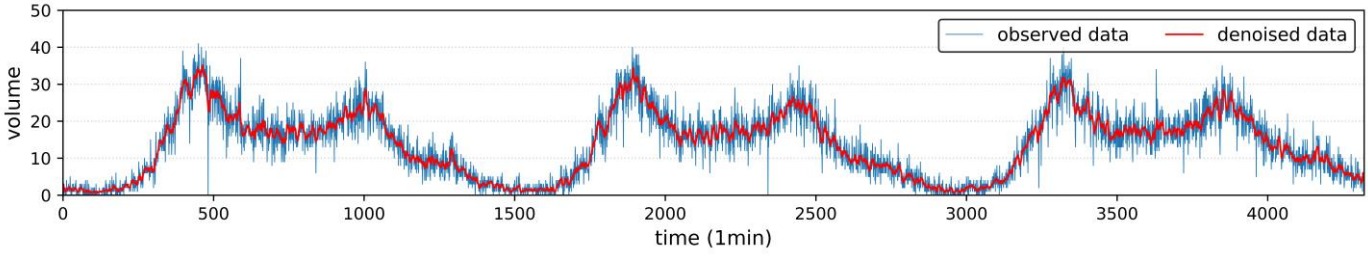
(e) denoising effect of MA3



(f) denoising effect of MA5

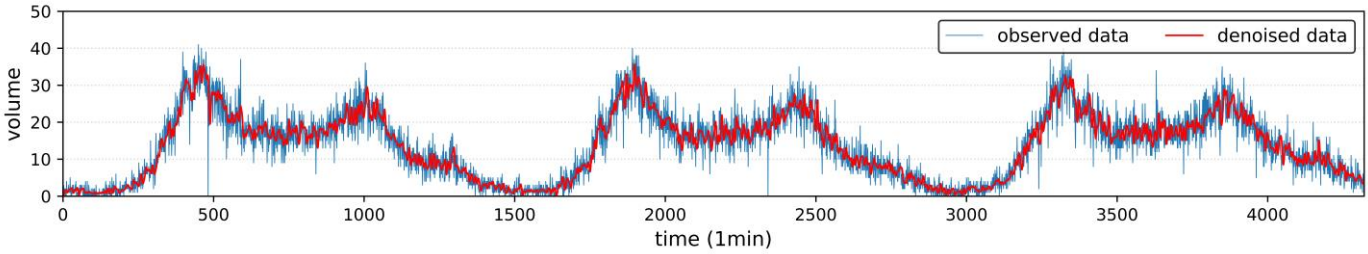


(g) denoising effect of MA7



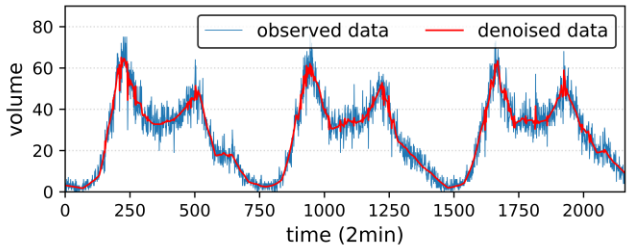
(h) denoising effect of MA9

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2020.3007809, IEEE Sensors Journal

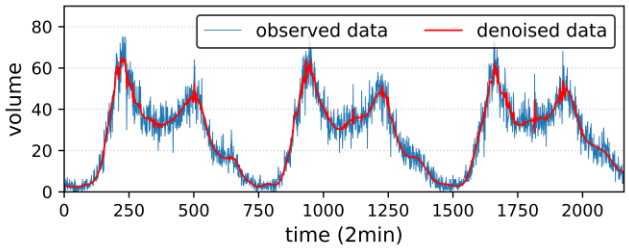


(i) denoising effect of BW

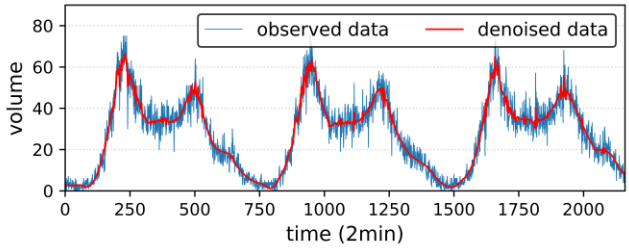
Fig. 2. Denoising effect of traffic flow data with sensor ID # 5802 under 1 min scale (note that time axis is the traffic flow data sample sequence)



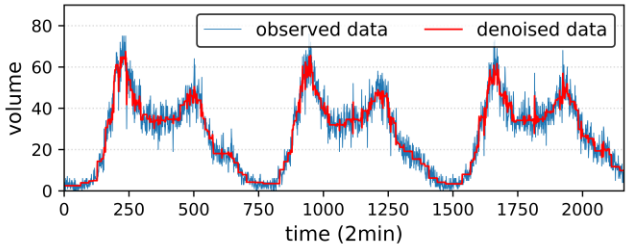
(a) denoising effect of WL (db)



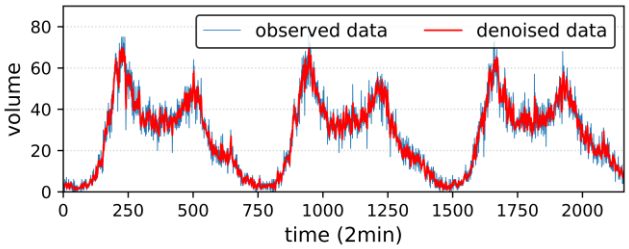
(b) denoising effect of WL (coif)



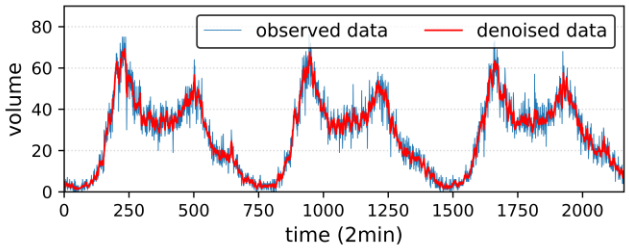
(c) denoising effect of WL (sym)



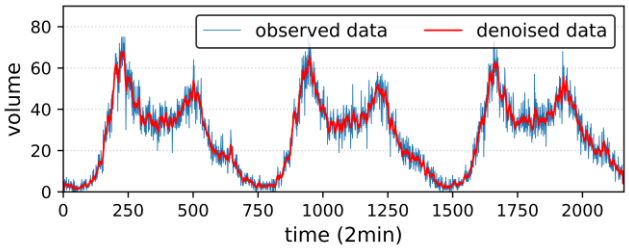
(d) denoising effect of WL (haar)



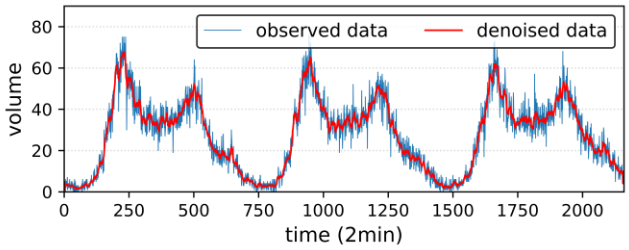
(e) denoising effect of MA3



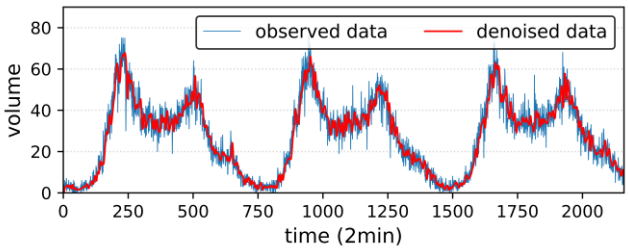
(f) denoising effect of MA5



(g) denoising effect of MA7

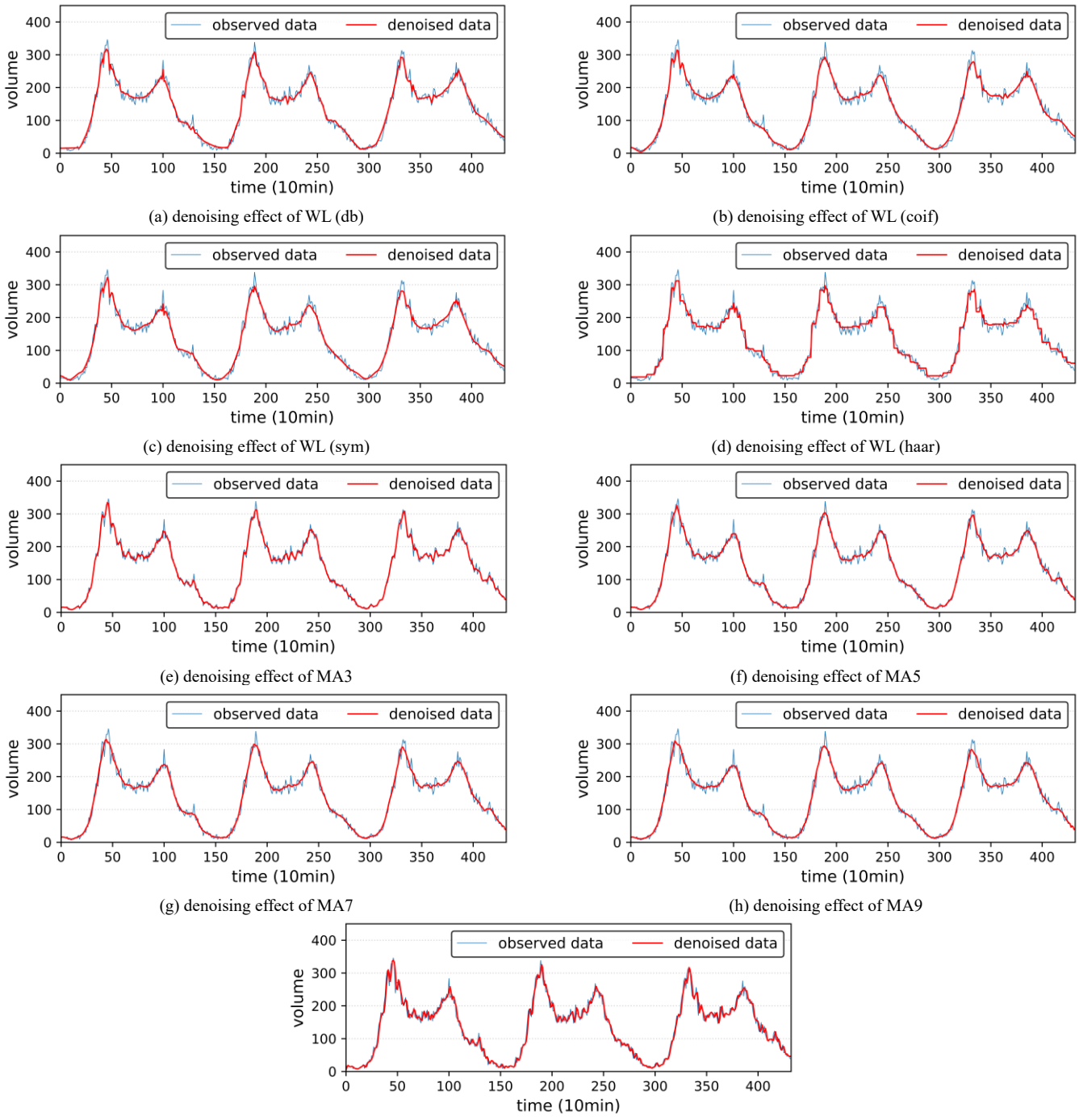


(h) denoising effect of MA9



(i) denoising effect of BW

Fig.3. Denoising effect of traffic flow data with sensor ID # 5802 under 2 min scale



(a) denoising effect of WL (db)

(b) denoising effect of WL (coif)

(c) denoising effect of WL (sym)

(d) denoising effect of WL (haar)

(e) denoising effect of MA3

(f) denoising effect of MA5

(g) denoising effect of MA7

(h) denoising effect of MA9

(i) denoising effect of BW

Fig.4. Denoising effect of traffic flow data with sensor ID # 5802 under 10 min scale

C. Prediction Performance Analysis

After removing anomalies in traffic flow data, we employ the ANN model to predict traffic flow volume at different time intervals. Note that 70% samples were selected as the training dataset, and the remaining 30% as the validation dataset. It is observed that the conventional ANN model obtains the largest error in comparison with hybrid ANN models (with data smoothing procedure). From the perspective of RMSE for 1 min (see Table 4), the conventional ANN obtained RMSE is 3.8977, which is obviously larger than the counterparts of smooth-framework based ANN models. We observed that the hybrid ANN model with BW denoising method showed better prediction accuracy (i.e., the RMSE, MAE and MAPE are 0.2755, 0.2094 and 0.0241) in comparison with other hybrid ANN models. The statistic indicators for 2 min and 10 min prediction accuracy showed quite similar results. The main reason is that samples near outliers (e.g., peak, spike, dips) were better smoothed into by the BW model, which benefits traffic flow prediction. Moreover, the prediction accuracy for 2 min

data is lower than that of the 1 min, while 10 min is worse than the 2 min (see Table 5 and 6), which indicates that larger time span can impose negative effect on traffic flow prediction.

We found that the prediction results (RMSE, MAE and MAPE) become worse when data sampling with a larger time interval (show Fig. 5). The main reason can be ascribed to randomness interference when aggregating traffic flow data into larger time interval. More specifically, traffic flow data in 10 min scale may contain unexpected oscillations which cannot be fully exploited by prediction models. Additionally, we found that the hybrid ANN models with WL models outperforms than that of MA-relevant models. The reason is that traffic flow data at different time scales obtains similar variation tendency, which can be better fitted by the WL models. Thus, the WL-based hybrid ANN models can extract more intrinsic traffic flow data pattern, and outputs better prediction results. Fig. 5 indicates that hybrid ANN combining with MA models showed better performance when the window size applied to MA models becomes larger.

Table 4. Prediction accuracy measurements for traffic data with sensor ID #5802

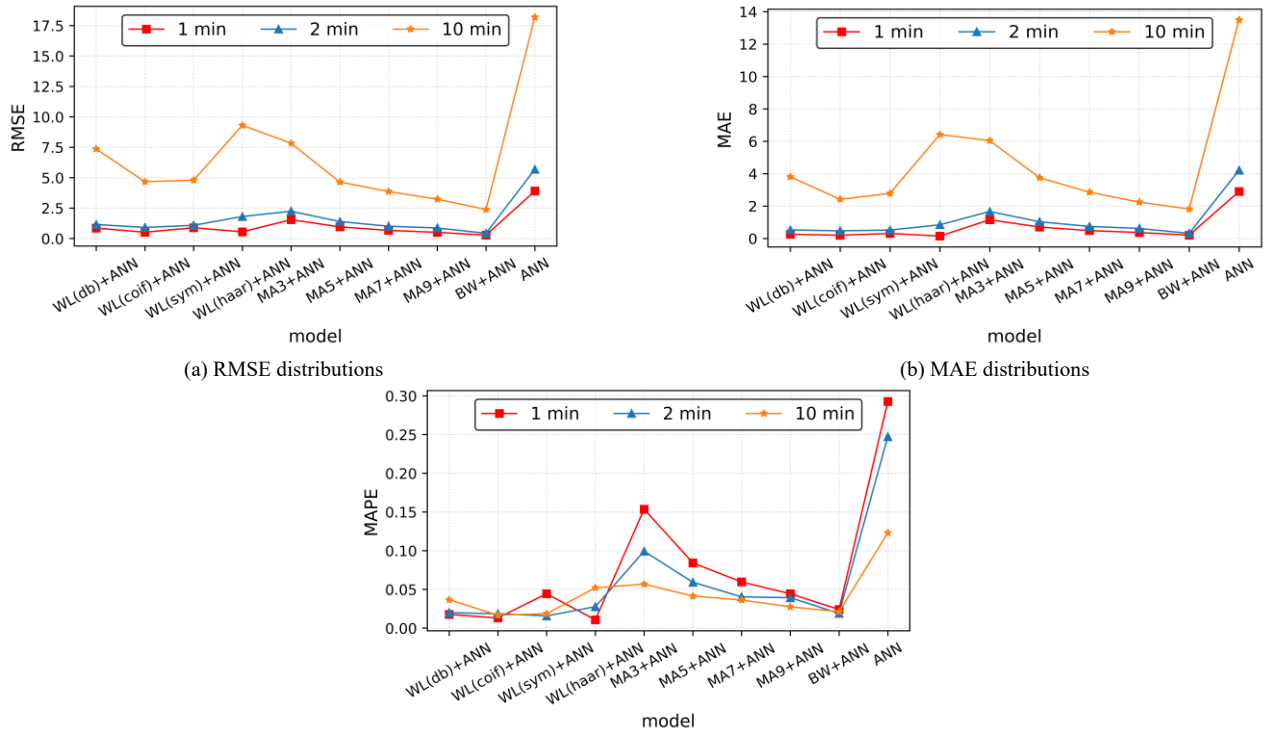
	1 min			2 min			10 min		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
WL (db)+ANN	0.8549	0.2624	0.0177	1.1482	0.5311	0.0198	7.3590	3.8150	0.0367
WL (coif)+ANN	0.5132	0.2007	0.0131	0.9060	0.4766	0.0185	4.6657	2.4198	0.0166
WL (sym)+ANN	0.8798	0.3148	0.0444	1.0833	0.5227	0.0157	4.7877	2.7885	0.0186
WL (haar)+ANN	0.5468	0.1502	0.0106	1.8045	0.8499	0.0276	9.2959	6.4195	0.0520
MA3+ANN	1.5419	1.1595	0.1536	2.2372	1.6704	0.0993	7.8253	6.0409	0.0568
MA5+ANN	0.9560	0.7099	0.0842	1.3913	1.0360	0.0592	4.6388	3.7540	0.0414
MA7+ANN	0.6629	0.4897	0.0596	1.0053	0.7480	0.0403	3.8578	2.8589	0.0362
MA9+ANN	0.5018	0.3734	0.0444	0.8638	0.6241	0.0392	3.2269	2.2481	0.0274
BW+ANN	0.2755	0.2094	0.0241	0.4244	0.3200	0.0190	2.3824	1.8221	0.0212
ANN	3.8977	2.9065	0.2927	5.6794	4.2151	0.2472	18.1861	13.4901	0.1231

Table 5. Prediction accuracy measurements for traffic data with sensor ID #5805

	1 min			2 min			10 min		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
WL (db)+ANN	0.5461	0.2190	0.0152	1.2585	0.5755	0.0223	10.1114	5.5282	0.0329
WL (coif)+ANN	0.7022	0.2721	0.0180	1.3270	0.5053	0.0165	11.2414	4.2668	0.0294
WL (sym)+ANN	0.5847	0.2351	0.0147	1.3483	0.6012	0.0165	10.7525	7.0071	0.0394
WL (haar)+ANN	0.7078	0.2103	0.0166	1.8903	0.5708	0.0226	16.0846	7.9766	0.0724
MA3+ANN	1.5923	1.1824	0.1366	2.3342	1.7315	0.1059	9.0346	6.6456	0.0792
MA5+ANN	0.9457	0.7127	0.0815	1.4471	1.0574	0.0604	5.3180	4.0465	0.0405
MA7+ANN	0.7375	0.5433	0.0615	0.9842	0.7354	0.0444	4.7638	3.3002	0.0335
MA9+ANN	0.5855	0.4362	0.0483	0.8449	0.6271	0.0373	3.1797	2.5027	0.0369
BW+ANN	0.2871	0.2126	0.0238	0.5035	0.3554	0.0192	2.3249	1.8023	0.0249
ANN	4.0356	3.0493	0.3164	5.9500	4.5087	0.2594	23.5575	16.8514	0.1338

Table 6. Prediction accuracy measurements for traffic data with sensor ID #5808

	1 min			2 min			10 min		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
WL (db)+ANN	0.5597	0.2161	0.0159	1.0847	0.5285	0.0232	6.4790	3.4729	0.0277
WL (coif)+ANN	0.3812	0.1579	0.0113	0.7469	0.3593	0.0118	4.5368	2.6163	0.0262
WL (sym)+ANN	0.5790	0.2058	0.0144	0.8552	0.4340	0.0183	5.1320	3.0364	0.0221
WL (haar)+ANN	0.3914	0.1203	0.0128	1.2806	0.3486	0.0146	12.4921	6.1300	0.0626
MA3+ANN	1.5114	1.1320	0.1493	2.0613	1.5500	0.0947	6.9239	5.3064	0.0636
MA5+ANN	0.8642	0.6538	0.0829	1.3544	1.0294	0.0578	3.7080	2.8675	0.0407
MA7+ANN	0.6576	0.4963	0.0585	1.0165	0.7525	0.0428	3.6047	2.7541	0.0406
MA9+ANN	0.5125	0.3756	0.0457	0.7798	0.5786	0.0344	3.0184	2.3250	0.0297
BW+ANN	0.2716	0.2071	0.0252	0.3985	0.3031	0.0197	2.4588	1.8147	0.0215
ANN	3.7075	2.7826	0.3369	5.3794	4.1256	0.2627	18.6181	13.4565	0.1485



(a) RMSE distributions

(b) MAE distributions

(c) MAPE distributions

Fig.5 Prediction accuracy distributions with time scales for sensor ID# 5802

IV. Conclusion

Short term traffic flow data provides crucial on-spot traffic state information for traffic participants (e.g., drivers on roadways, traffic management officials), and thus obtaining accurate traffic flow data in advance attracts significant attentions in the traffic flow community. We proposed a novel framework for the purpose of accurately predicting short-term traffic flow (i.e., 1 min, 2 min and 10 min) from historical data. We firstly denoised the raw traffic flow data with popular smoothing models (WL, MA and BW). After that, the ANN model was introduced to forecast traffic flow based on the smoothed traffic flow data. We have evaluated the proposed model performance on the empirical traffic flow data collected from three loop sensors. The experimental results support the following conclusions: (1) traffic flow prediction accuracy obtained by hybrid ANN model combined with denoising methods is superior to that of the non-hybrid ANN model; (2) the MA based hybrid ANN models showed an increasing prediction accuracy when the window size becomes larger. In other words, larger window size used in the MA model can benefit prediction accuracy; (3) the hybrid ANN model combined with WL denoising methods are superior than the MA relevant counterpart when implementing the traffic flow prediction accuracy; (4) the Butterworth filter supported hybrid ANN model obtained better prediction accuracy compared to the counterparts. The study was implemented without considering environmental factor influence, which may affect the prediction error. We can expand our research by incorporating various external factors' influence (e.g., roadway maintenance, adverse weather, unexpected sports event, etc.) in a quantitative manner. Besides, we can explore lightweight deep learning model performance to implement the short term traffic flow prediction task. Moreover, we can testify the WL filter smoothing performance at different decomposition levels for the purpose of obtaining more holistic comparison results.

References

[1] H.-F. Yang, T. S. Dillon, E. Chang, and Y.-P. P. Chen, "Optimized configuration of exponential smoothing and extreme learning machine for traffic flow forecasting," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 23-34, 2018.

[2] M. Kontorinaki, I. Karafyllis, and M. Papageorgiou, "Global exponential stabilisation of acyclic traffic networks," *International Journal of Control*, vol. 92, pp. 564-584, 2019.

[3] L. Cai, Z. Zhang, J. Yang, Y. Yu, T. Zhou, and J. Qin, "A noise-immune Kalman filter for short-term traffic flow forecasting," *Physica A: Statistical Mechanics and its Applications*, vol. 536, p. 122601, 2019/12/15/ 2019.

[4] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 50-64, 2014.

[5] D. Ming-Jun and Q. Shi-Ru, "Fuzzy state transition and Kalman filter applied in short-term traffic flow forecasting," *Computational intelligence and neuroscience*, vol. 2015, 2015.

[6] A. Anwar and S. Duraisamy, "A Predictive Routing Algorithm for WBSN Based on Kalman Filter Iterations," *IEEE Sensors Journal*, vol. 18, pp. 7741-7748, 2018.

[7] P. Duan, G. Mao, W. Liang, and D. Zhang, "A unified spatio-temporal model for short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 3212-3223, 2018.

[8] Q. Hou, J. Leng, G. Ma, W. Liu, and Y. Cheng, "An adaptive hybrid model for short-term urban traffic flow prediction," *Physica A: Statistical Mechanics and its Applications*, vol. 527, p. 121065, 2019.

[9] K.-L. Li, C.-J. Zhai, and J.-M. Xu, "Short-term traffic flow prediction using a methodology based on ARIMA and RBF-ANN," in *2017 Chinese Automation Congress (CAC)*, 2017, pp. 2804-2807.

[10] C. Xu, Z. Li, and W. Wang, "Short-term traffic flow prediction using a methodology based on autoregressive integrated moving average and genetic programming," *Transport*, vol. 31, pp. 343-358, 2016.

[11] H. El-Sayed, S. Sankar, Y.-A. Daraghmi, P. Tiwari, E. Rattagan, M. Mohanty, *et al.*, "Accurate traffic flow prediction in heterogeneous vehicular networks in an intelligent transport system using a supervised non-parametric classifier," *Sensors*, vol. 18, p. 1696, 2018.

[12] J. Tang, X. Chen, Z. Hu, F. Zong, C. Han, and L. Li, "Traffic flow prediction based on combination of support vector machine and data denoising schemes," *Physica A: Statistical Mechanics and its Applications*, vol. 534, p. 120642, 2019.

[13] P. Vijayaraju, B. Sripathy, D. Arivudainambi, and S. Balaji, "Hybrid Memetic Algorithm With Two-Dimensional Discrete Haar Wavelet Transform for Optimal Sensor Placement," *IEEE Sensors Journal*, vol. 17, pp. 2267-2278, 2017.

[14] S. Goudarzi, M. N. Kama, M. H. Anisi, S. A. Soleymani, and F. Doctor, "Self-organizing traffic flow prediction with an optimized deep belief network for internet of vehicles," *Sensors*, vol. 18, p. 3459, 2018.

[15] D. Zhu, H. Du, Y. Sun, and N. Cao, "Research on path planning model based on short-term traffic flow prediction in intelligent transportation system," *Sensors*, vol. 18, p. 4275, 2018.

[16] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transportation Research Part C: Emerging Technologies*, vol. 54, pp. 187-197, 2015/05/01/ 2015.

[17] F. Moretti, S. Pizzuti, S. Panzieri, and M. Annunziato, "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling," *Neurocomputing*, vol. 167, pp. 3-7, 2015.

[18] J. Tang, J. Hu, W. Hao, X. Chen, and Y. Qi, "Markov Chains based route travel time estimation considering link spatio-temporal correlation," *Physica A: Statistical Mechanics and its Applications*, p. 123759, 2019.

[19] Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, pp. 865-873, 2015.

[20] M. Khajeh Hosseini and A. Talebpour, "Traffic Prediction using Time-Space Diagram: A Convolutional Neural Network Approach," *Transportation Research Record*, vol. 2673, pp. 425-435, 2019.

[21] D. Kim and O. Jeong, "Cooperative traffic signal control with traffic flow prediction in multi-intersection," *Sensors*, vol. 20, p. 137, 2020.

[22] N. Lu, Y. Ma, and Y. Liu, "Evaluating probabilistic traffic load effects on large bridges using long-term traffic monitoring data," *Sensors*, vol. 19, p. 5056, 2019.

[23] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction," *Sensors*, vol. 17, 2017.

[24] X. Chen, Y. Yang, S. Wang, H. Wu, J. Tang, J. Zhao, *et al.*, "Ship Type Recognition via a Coarse-to-Fine Cascaded Convolution Neural Network," *Journal of Navigation*, vol. 73, pp. 813-832, 2020.

[25] X. Yang, Y. Zou, J. Tang, J. Liang, and M. Ijaz, "Evaluation of Short-Term Freeway Speed Prediction Based on Periodic Analysis Using Statistical Models and Machine Learning Models," *Journal of Advanced Transportation*, vol. 2020, pp. 1-16, 01/20 2020.

[26] X. Chen, L. Qi, Y. Yang, Q. Luo, O. Postolache, J. Tang, *et al.*, "Video-Based Detection Infrastructure Enhancement for Automated Ship Recognition and Behavior Analysis," *Journal of Advanced Transportation*, vol. 2020, pp. 1-12, 2020.

[27] X. Chen, H. Chen, H. Wu, Y. Huang, Y. Yang, W. Zhang, *et al.*, "Robust Visual Ship Tracking with an Ensemble Framework via Multi-View Learning and Wavelet Filter," *Sensors (Basel)*, vol. 20, Feb 10 2020.

[28] J. Tang, F. Gao, F. Liu, and X. Chen, "A Denoising Scheme-Based Traffic Flow Prediction Model: Combination of Ensemble Empirical Mode Decomposition and Fuzzy C-Means Neural Network," *IEEE Access*, vol. 8, pp. 11546-11559, 2020.

[29] X. Chen, Z. Li, Y. Wang, J. Tang, W. Zhu, C. Shi, *et al.*, "Anomaly Detection and Cleaning of Highway Elevation Data from Google Earth Using Ensemble Empirical Mode Decomposition," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, p. 04018015, 2018.

[30] Y. Peng and W. Xiang, "Short-term traffic volume prediction using GA-BP based on wavelet denoising and phase space reconstruction," *Physica A: Statistical Mechanics and its Applications*, p. 123913, 2019.

[31] P. Sun, N. AlJeri, and A. Boukerche, "A fast vehicular traffic flow prediction scheme based on fourier and wavelet analysis," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1-6.

[32] F. G. Habtemichael and M. Cetin, "Short-term traffic flow rate forecasting based on identifying similar traffic patterns," *Transportation research Part C: emerging technologies*, vol. 66, pp. 61-78, 2016.

[33] D. Xia, B. Wang, H. Li, Y. Li, and Z. Zhang, "A distributed spatial–temporal weighted model on MapReduce for short-term traffic flow forecasting," *Neurocomputing*, vol. 179, pp. 246-263, 2016.

[34] S. Zhang, Y. Yao, J. Hu, Y. Zhao, S. Li, and J. Hu, "Deep autoencoder neural networks for short-term traffic congestion prediction of transportation networks," *Sensors*, vol. 19, p. 2229, 2019.

[35] J. Abdi, B. Moshiri, B. Abdulhai, and A. K. Sedigh, "Forecasting of short-term traffic-flow based on improved neurofuzzy models via emotional temporal difference learning algorithm," *Engineering Applications of Artificial Intelligence*, vol. 25, pp. 1022-1042, 2012.

[36] X. Luo, D. Li, Y. Yang, and S. Zhang, "Spatiotemporal traffic flow prediction with KNN and LSTM," *Journal of Advanced Transportation*, vol. 2019, 2019.

[37] X. Luo, L. Niu, and S. Zhang, "An algorithm for traffic flow prediction based on improved SARIMA and GA," *KSCE Journal of Civil Engineering*, vol. 22, pp. 4107-4115, 2018.

[38] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, p. 1501, 2017.

[39] Q. Luo, J. Yuan, X. Chen, J. Yang, W. Zhang, and J. Zhao, "Research on Mixed User Equilibrium Model Based on Mobile Internet Traffic Information Service," *IEEE Access*, vol. 7, pp. 164775-164791, 2019.

[40] X. Chen, S. Wang, C. Shi, H. Wu, J. Zhao, and J. Fu, "Robust Ship Tracking via Multi-view Learning and Sparse Representation," *Journal of Navigation*, vol. 72, pp. 176-192, 2019.

[41] X. Chen, X. Xu, Y. Yang, H. Wu, J. Tang, and J. Zhao, "Augmented Ship Tracking Under Occlusion Conditions From Maritime Surveillance Videos," *IEEE Access*, vol. 8, pp. 42884-42897, 2020.

[42] X. Chen, J. Lu, J. Zhao, Z. Qu, Y. Yang, and J. Xian, "Traffic Flow Prediction at Varied Time Scales via Ensemble Empirical Mode Decomposition and Artificial Neural Network," *Sustainability*, vol. 12, pp. 1-17, 2020.

[43] Y. Rajabzadeh, A. H. Rezaie, and H. Amindavar, "Short-term traffic flow prediction using time-varying Vasicek model," *Transportation Research Part C: Emerging Technologies*, vol. 74, pp. 168-181, 2017/01/01/ 2017.

[44] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong, "An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction," *Procedia - Social and Behavioral Sciences*, vol. 96, pp. 653-662, 2013/11/06/ 2013.

[45] D.-w. Huang, "Wavelet analysis in a traffic model," *Physica A: Statistical Mechanics and its Applications*, vol. 329, pp. 298-308, 2003.

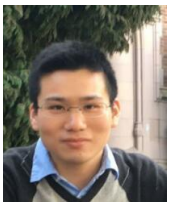
[46] S. Jangjit and M. Ketcham, "A New Wavelet Denoising Method for Noise Threshold," *Engineering Journal*, vol. 21, pp. 141-155, 2017.

[47] O. Singh and R. K. Sunkaria, "ECG signal denoising based on empirical mode decomposition and moving average filter," in *2013 IEEE International Conference on Signal Processing, Computing and Control (ISPCC)*, 2013, pp. 1-6.

[48] R. Ezzeldin and A. Hatata, "Application of NARX neural network model for discharge prediction through lateral orifices," *Alexandria engineering journal*, vol. 57, pp. 2991-2998, 2018.

[49] J. M. P. Menezes Jr and G. A. Barreto, "Long-term time series prediction with the NARX network: An empirical evaluation," *Neurocomputing*, vol. 71, pp. 3335-3343, 2008.

[50] E. Pisoni, M. Farina, C. Carnevale, and L. Piroddi, "Forecasting peak air pollution levels using NARX models," *Engineering Applications of Artificial Intelligence*, vol. 22, pp. 593-602, 2009.



Xinqiang Chen received his Ph.D. degree (2018) in traffic information engineering and controlling at the Shanghai Maritime University in China. From September 2015 to September 2016, he was a visiting student in the Smart Transportation Applications and Research Laboratory at the University of Washington, United States. He is the author and coauthor of more than 13 technical papers. His research interests include traffic data analysis, transportation image processing, intelligent transportation systems, transportation video analysis, smart ship etc.

Shubo Wu received bachelor degree in traffic engineering in Guangzhou University, Guangzhou, China, in 2018. And now he is a master student in Shanghai Maritime University, Shanghai, China with major of traffic information engineering and control. His research interests are traffic big data analysis and data mining.

Chaojian Shi graduated from Shanghai Maritime University, Shanghai China, in 1982. From 1986 to 1987, He received further education majoring in advanced navigation science. He served as many academic memberships for many journals and conference. His research field includes but not limited to the following lists: artificial intelligence, image analysis and processing, computer vision, machine learning and so on.

Yanguo Huang serves as a professor at School of electrical engineering and automation at Jiangxi University of Science and technology. He has author and coauthored more than 20 papers. His research interests include intelligent transportation systems, electrical vehicle modeling, etc.

Yongsheng Yang received his Ph.D. (1998) degree from Nanjing University of Aeronautics and Astronautics in China. He is currently a professor at the Shanghai Maritime University. He serves as associate editor for the journal of Computer Aided Engineering. He is the author and coauthor of more than 50 technical papers and proceedings. His research interests include operation and optimization of port logistics, cooperated job scheduling and controlling for automatic terminals etc.

Ruimin Ke (Member, IEEE) received his B.E. degree from the Department of Automation at Tsinghua University in 2014, and the M.S. degree from the Civil and Environmental Engineering at University of Washington in 2016. He is currently working toward the Ph.D. degree in Civil and Environmental Engineering at University of Washington. His research interests include intelligent transportation systems, transportation data science, smart city, autonomous driving, internet of things, and computer vision. He is a member of the Statewide and National Data and Information Management Committee of the TRB, a young member of the Infrastructure Systems Committee of ASCE T&DI, and a member of the Urban Computing and Space Optimization Committee of WTC.

Jiansen Zhao received the Ph.D. degree in marine engineering in Dalian Maritime University, Dalian, China, in 2013. And now he is a senior lecturer in Shanghai Maritime University, Shanghai, China. His current research interests include intelligent navigation, plasma antenna, gas discharge, AIS antenna, and GNSS antenna.