IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 4, 2018, accepted September 27, 2018, date of publication October 5, 2018,
date of current version October 25, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2874068

Vehicle Emission Forecasting Based on Wavelet Transform and Long Short-Term Memory Network

QIANG ZHANG[1], FENG LI[1], FEI LONG[2], AND QIANG LING[1], (Senior Member, IEEE)
[1]Department of Automation, University of Science and Technology of China, Hefei 230027, China
[2]Chinaso, Inc., Beijing 100077, China

Corresponding author: Qiang Ling (qling@ustc.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2016YFC0201003 and in part by the Internet plus major project for the Internet plus coordinated manufacturing cloud service support platform.

ABSTRACT This paper proposes a time series model based on wavelet transform and long short-term memory (**LSTM**) network to forecast vehicle emission. It implements the semi-supervised collaborative training regression to compensate missing emissions data. The accumulated carbon monoxide (CO), hydrocarbons (HC), and nitric oxide (NO) concentrations emitted by vehicles in different lanes per hour were taken to quantitatively characterize the vehicle emissions. The original time series of vehicle emission data, which may be highly variable, is decomposed into several lowly variable sub-series by wavelet transform. For each sub-series, an LSTM time series model is proposed to forecast vehicle emissions. More specifically, the inputs of that LSTM model are the weather variables, the driving variables of the concerned vehicle and historical emissions records while its output is the predicted accumulated concentrations of CO, HC, and NO. The three types of predicted concentrations of all sub-series are summed up, respectively, and produce the desired prediction of the total emission of each type. The proposed model is verified through real data which was collected between May 2017 and December 2017 at the multi-lane monitoring station of Baimiao South Road, Daxing District, Beijing, China. It confirms that our model based on wavelet transform and LSTM can efficiently improve the correlation coefficient (R) and the index of agreement (IA) against conventional models, such as ARIMA and wavelet-ARIMA model.

INDEX TERMS Vehicle emissions forecasting, long short-term memory network, wavelet transform, semi-supervised collaborative training regression.

I. INTRODUCTION

Vehicle emission causes poor air quality and may induce various respiratory diseases. So vehicle emission monitoring and control is important for improving air quality. The current vehicle emission measure methods mainly include the chassis and engine dynamometer testing, tunnel measure method, portable device detection method, and remote sensing detection. The chassis and engine dynamometer testing measure is only suitable for specific simulated driving conditions [1]. The tunnel method measures the air flow through a tunnel based on the fact that the difference of gas concentrations between the outlet and the inlet can represent the emissions of vehicles, and may be sensitive to meteorological conditions [2]. A portable device, such as a PEMS system, connects the probe to the exhaust pipe of the vehicle and can precisely measure the emissions, but, at a high cost [3]. The remote sensing detection is a non-contact on-road vehicle emission measure method, which has the advantages of real-time detection, low cost, and no disturbance to the normally running vehicles [4] [5]. Therefore, the emission data in our study was collected by the remote sensing measure method.

From the perspective of air quality monitoring and emission control, the accumulated vehicle emissions in a given time duration, e.g., one hour, are more meaningful than the emissions of a single vehicle and are, therefore, considered in our study. Furthermore, we attempt to forecast the future accumulated emissions, e.g., the ones in a few hours, with the current and historical results. As vehicle emissions are affects by geographical conditions, traffic flow, meteorological

conditions and other factors, these factors are also taken as inputs of the emission forecasting.

Emission forecasting has caught much attention in the literature. Common emission forecasting methods mainly include Autoregressive Integrated Moving Average model (ARIMA), BP neural network, and Recurrent Neural Network(RNN). Cai [6] used an ARIMA model to analyze the time series of the largest hourly carbon monoxide (CO) concentration per month in the south coast of California and can yield accurate prediction. Kukkonen *et al.*[7] used multiple BP neural network models to predict the concentration of NO_2 and PM_{10} in Helsinki, Finland. Cai *et al.* [8] implemented artificial neural networks to predict hourly air pollutant concentrations near arteries in Guangzhou. Connor and Atlas [9] demonstrated that RNN performs better than general feed-forward neural networks when predicting vehicle emissions. Brunelli *et al.* [10] studied the applicability of recurrent neural networks (Elman model) for predicting the maximum concentration of daily SO_2, O_3, PM_{10}, NO_2 and other pollutants. Alizadeh *et al.* [11] performed the suspended sediment forecasting using wavelet-ANN ensemble models, and also investigated the effectiveness of wavelet-ANN methods for predicting rainfall and runoff in Tolt River basin [12]. Salazar *et al.* [13] proposed a new method to predict the ozone concentrations based on wavelet-ARIMA hybrid models.

In real applications, the ARIMA method requires the stationarity of the time series of emission data, which is not easy to guarantee, and may be sensitive to model parameters. As BP neural networks have no memory, they are not very effective in dealing with the time series of emission data whose future value may depend on both the current and the historic values. Although RNN usually yields higher prediction accuracy in dealing with the time series of emission data, it is prone to the gradient vanishing problem and suffers from short memory span. Hochreiter and Schmidhuber [14] showed that LSTM is a special kind of recurrent neural network and suitable for dealing with long-term dependence problems. Compared with the general recurrent neural networks, LSTM adds a cell state that preserves long-term information in the hidden layer and is more effective to avoid gradient vanishing problem and gradient exploding problem in dealing with long sequence problems. Therefore, our study proposes to implement LSTM to forecast vehicle emissions through analyzing the time series of vehicle emission data.

The original time series of vehicle emission data is usually highly variable, which may inhibit the prediction accuracy of the accumulated vehicle emissions. To resolve this issue, we implement the wavelet transform to decompose the highly variable time series of vehicle emission data into multiple lowly variable sub-series as [15]. For each sub-series, the LSTM method is implemented to forecast the accumulated concentrations of CO, HC and NO. Then the three types of predicted concentrations of all sub-series are summed up, respectively, and produce the desired prediction of the total emission of each type. We take some real data to verify the proposed method. The experimental results confirm that our methods based on wavelet transform and LSTM can achieve better forecasting performance than some state-of-the-art methods, such as ARIMA and wavelet-ARIMA.

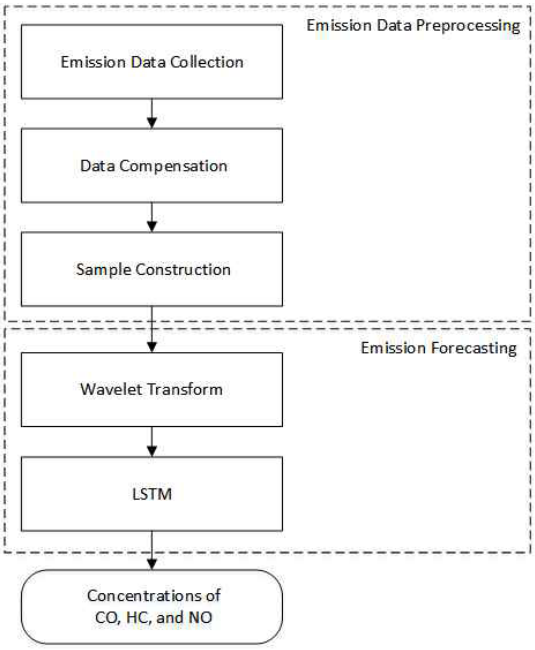

FIGURE 1. The overview of the emission forecasting model.

The remainder of the this paper is organized as follows. Section II presents the details of the proposed method. Experimental results are illustrated and analyzed in Section III. Some final remarks are placed in Section IV.

II. DATA AND METHODS

The major procedure of the proposed emission forecasting model is shown in Fig. 1. Based on the collected emission data, the concentrations of CO, HC and NO are forecasted. The related operations are introduced in the subsequent subsections.

A. DATA COLLECTION

The vehicle emission data investigated in this paper was collected by the Urban Road Network Motor Vehicle Emission Monitoring System from May to December 2017 at a multi-lane monitoring station of Baimiao South Road, Daxing District, Beijing, China. The meteorological data used in this paper was obtained from the China Meteorological Administration, including the hourly temperature, wind speed, wind direction, humidity and atmospheric pressure. The Urban Road Network Motor Vehicle Emission Monitoring System is mainly comprised of the remote sensing equipment host, the retroreflective sheeting, the license plate recognition system and the LED display screen as shown in Figure 2. This monitoring system works mainly based on the spectral feature absorption principle and video image analysis technology. It is a non-contact real-time emission

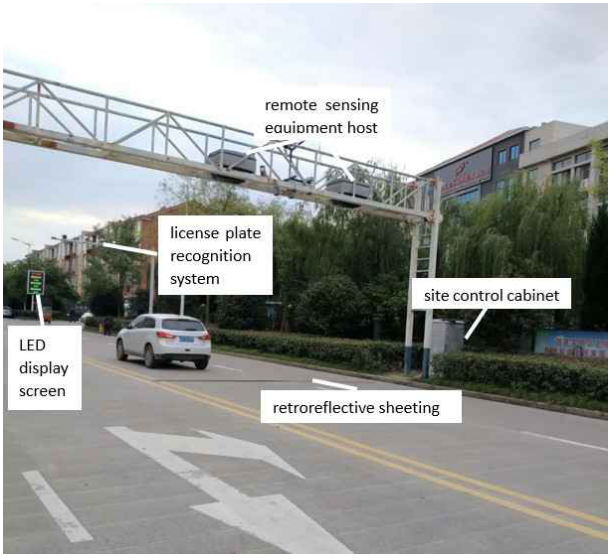

FIGURE 2. The Urban Road Network Vehicle Exhaust Emission Monitoring System to collect emission data.

detection method. After the exhaust of a vehicle is emitted, it will diffuse to form a plume and its opacity can be measured by the intensity change of the light after passing by the plume. The volume concentration ratio of each emission component in the plume is almost constant. Taking carbon dioxide (CO_2) as the reference emission gas, the volume concentration ratio of CO, HC and NO to CO_2 can be measured. Specifically, the ratio of CO to CO_2 in the emission gas is measured by the laser absorption spectroscopy in the near-infrared region, and the volume concentration ratios of HC and NO to CO_2 are measured by the differential spectrum of ultraviolet light. The volumetric concentration of CO, HC and NO in the emission gas is obtained by the inversion using the combustion chemical equation in (1), where Q is the volumetric concentration ratio of CO to CO_2, Q′ is the ratio of volumetric concentration of HC to CO_2, and Q″ is the ratio of volumetric concentration of NO to CO_2 [16].

$$\begin{aligned} \%CO_2 &= \frac{42}{2.79 + 2Q + 1.21Q' + Q''} \\ \%CO &= \%CO_2 \times Q \\ \%HC &= \%CO_2 \times Q' \\ \%NO &= \%CO_2 \times Q'' \end{aligned} \tag{1}$$

According to [16], the vehicle emission mass concentration can be expressed as

$$\begin{aligned} EF_{CO}(g \cdot km^{-1}) &= 1020.48 \frac{Q \cdot V^{-0.674}}{1 + Q + 3Q'/0.493} \\ EF_{HC}(g \cdot km^{-1}) &= 3105.27 \frac{Q' \cdot V^{-0.674}}{1 + Q + 3Q'/0.493} \\ EF_{NO}(g \cdot km^{-1}) &= 1099.57 \frac{Q'' \cdot V^{-0.674}}{1 + Q + 3Q'/0.493}, \end{aligned} \tag{2}$$

where V is the speed of the concerned vehicle. The license plate recognition subsystem can identify the license plate number of the concerned vehicle on the road through the fixed camera, video capture card and license plate recognition module. In addition, the integrated speedometer in the system measures the speed, acceleration and length of the concerned vehicle. The LED display shows the license plate number, speed, emission detection results of the concerned vehicle. The site control cabinet is a control unit with a database and a correlation analysis center.

B. DATA COMPENSATION

Some of the emission data recorded by the monitoring system is incomplete due to sensor failures or partial data loss during data transmission and updating. Therefore, before analyzing the emission data, it is necessary to compensate the missing data. There are two main types of missing data.

- *The remote sensing emission data is missing.* All local vehicles regularly go to the vehicle management station for exhaust emission testing. When the remote sensing emission data of one vehicle is missing, its last emission test results saved at the vehicle management station can be used for compensation.
- *Other non-emission data, such as vehicle-length, speed, acceleration, are missing.* We implement the semi-supervised collaborative training regression [17] to compensate the missing data. This method make up the missing data by iteratively training two K-NearestNeighbor (KNN) regressors, whose details are presented below.

1) Using the complete data set L to generate two KNN regressors h_1 and h_2 based on distance divergence.
2) Randomly select 50 incomplete data points from the incomplete data set U. For each point $X \in U$, use the regressors h_1 and h_2 to generate estimated values Y_1 and Y_2, respectively. (X,Y_1) updates h_1 to h_1' while (X,Y_2) updates h_2 to h_2'. Find the k-nearest neighbors of point X, Ω. Then define two sums of squared error difference, e_1 and e_2, as

$$e_1 = \sum_{X_i \in \Omega} ((Y_i - h_1(X_i))^2 - (Y_i - h_1'(X_i))^2) \tag{3}$$

$$e_2 = \sum_{X_i \in \Omega} ((Y_i - h_2(X_i))^2 - (Y_i - h_2'(X_i))^2) \tag{4}$$

   Find the minimum e_1 and e_2 among the selected 50 points, which are denoted as e_1' and e_2', respectively. The found two points are denoted as $(X_1', h_1(X_1'))$, $(X_2', h_2(X_2'))$.
3) If $e_1' > 0$, remove the point $(X_1', h_1(X_1'))$ from the incomplete data set U, add it into the training set L_2 of h_2 and update h_2. If $e_2' > 0$, remove the point $(X_2', h_2(X_2'))$ from the incomplete data set U, and add it into the training set L_1 of h_1 and update h_1.
4) Repeat 2)–3) until there are no incomplete data points or less than 50 points in U. Then return the output regressor $h = 0.5(h_1 + h_2)$, and use h to generate the value of each incomplete point in U.

TABLE 1. The absolute value of Pearson correlation coefficients of input and output features.

$\|r\|$	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}
F_1	1	0.25	0.39	0.10	0.21	0.25	0.28	0.19	0.15	0.76	0.91	0.50
F_2	0.25	1	0.32	0.01	0.15	016	0.07	0.18	0.03	0.31	0.33	0.24
F_3	0.39	0.32	1	0.12	0.14	0.12	0.18	0.06	0.17	0.20	0.41	0.26
F_4	0.10	0.01	0.12	1	0.14	0.06	0.02	0.17	0.01	0.29	0.26	0.15
F_5	0.21	0.15	0.14	0.14	1	0.31	0.28	0.05	0.40	0.22	0.21	0.19
F_6	0.25	0.16	0.12	0.06	0.31	1	0.26	0.21	0.25	0.28	0.29	0.22
F_7	0.28	0.07	0.18	0.02	0.28	0.26	1	0.61	0.74	0.35	0.44	0.26
F_8	0.19	0.18	0.06	0.17	0.05	0.21	0.61	1	0.70	0.11	0.19	0.18
F_9	0.15	0.03	0.17	0.01	0.40	0.25	0.74	0.70	1	0.34	0.38	0.25
F_{10}	0.76	0.31	0.20	0.29	0.22	0.28	0.35	0.11	0.34	1	0.56	0.61
F_{11}	0.91	0.33	0.41	0.26	0.21	0.29	0.44	0.19	0.38	0.56	1	0.38
F_{12}	0.50	0.24	0.26	0.15	0.19	0.22	0.26	0.18	0.25	0.61	0.38	1

In fact, the above KNN regressors can also be be achieved by choosing different k-neighbor values. Based on divergence, KNN can prevent the two regressors from degenerating into the same regressor, and take advantage of the available information regarding incomplete data. Actually, it is a lazy algorithm. When we update the KNN Regressors, we do not update all data and just update some *necessary* points. For example, our experiments have 215,865 original remote sensing monitoring vehicle records at the multi-lane emission monitoring station of Baimiao South Road, Daxing District, Beijing,China, from May 1 to December 31, 2017. And there are 4,402 records without CO, HC, or NO emission concentration, and 7403 other incomplete. Through the above compensation method, a total of 11805 incomplete data records are compensated and fundamentally improves the forecasting performance.

C. CONSTRUCTION OF EMISSIONS DATA SAMPLES

After compensating the missing emission data by the method in Section II-B, the monitoring emission data records are complete. However, these original monitoring records characterize only the emissions of individual vehicles. In order to estimate and predict the overall vehicle emissions of different lanes at the concerned monitoring station, it is necessary to construct a time series of emission samples that can characterize the regional emission. We calculated the hourly total number of vehicles passing by the multi-lane monitoring station of Baimiao South Road in Daxing District, Beijing, China, during the period from May 1 to December 31, 2017. For these vehicles, their average length, average speed, average acceleration, accumulated hourly concentrations of CO, HC and NO are also computed. Moreover, the hourly temperature, wind, wind direction, relative humidity, atmospheric pressure are obtained from the meteorological department as the components of an hourly emission sample. In order to build an emission forecasting model more accurately, we delete the entire 24-hour records of the days when more than 3 hours' vehicle emission data are not recorded.

We use the accumulated hourly concentrations of CO,HC and NO as the overall reference estimated emissions. We use the emissions data samples in past 5 hours, including the number of vehicles, the average vehicle length, the average driving conditions,and the meteorological conditions, to forecast the accumulated hourly concentrations of CO,HC and NO in 1 or 3 hours. Here the average vehicle length can indicate the type of passing vehicles to a certain extent. Generally, the longer the length of a vehicle, higher emission it exhausts. Vehicle emissions are closely related to driving conditions. In the normal driving speed range, the slower the vehicle speed is, the higher emission we can expect [18]. In addition, vehicle emissions and plume diffusion also depend on meteorological conditions, such as temperature and wind speed. Due to the aforementioned reasons, we correspondingly construct the emission data samples and set the input and output of the emission forecasting model as above.

D. CORRELATION ANALYSIS

The input features of the time series model are expected to be highly correlated with the output while the multiple input features are expected to be lowly correlated with each other. We use the absolute value of Pearson correlation coefficient($|r|$) to represent the correlation between the input and output features. These correlation results are shown in Table 1, where $F_1, F_2, \cdots, F_9$ stand for 9 input features, including the hourly number of vehicles passing by the monitoring lanes, the average vehicle length, speed, acceleration of these vehicles, the wind speed and direction, the outdoor temperature, the relative humidity, the atmospheric pressure, and F_{10}, F_{11}, F_{12} are the 3 output features, including the accumulated hourly concentrations of CO, HC and NO. Of course, the autocorrelation coefficient of each feature in Table 1 is 1. The correlation coefficients between the the input features from F_1 to F_9 are mainly small, which are often less than 0.25 and have an average of 0.215. The correlation coefficients between the output features (from F_{10} to F_{12}) and the input features (from F_1 to F_9) are larger and have an average of 0.320. By Table 1, we see that there is a strong correlation between F_1 and the three output features, which results from the fact that more vehicles (larger F_1) produce more emissions, i.e., larger accumulated hourly

concentrations of CO, HC and NO. Table 1 also demonstrates a strong correlation between the meteorological conditions of the input features, such as F_7 (the temperature), F_8 (the relative humidity) and F_9 (the atmospheric pressure).

E. WAVELET TRANSFORM

The original emissions time series constructed with the above emission samples are usually highly variable and difficult to analyze and predict. So we introduce the wavelet transform to decompose the original highly variable time series into several lowly variable sub-series. Then, each sub-series is processed by a LSTM network to forecast the accumulated hourly concentrations of CO, HC and NO. Finally, the forecasted concentrations from these sub-series are summed up to yield the desired total concentrations. In this forecasting procedure, the LSTM network plays a critical role and its details are explained below.

Compared with discrete Fourier transform, wavelet transform can not only obtain the frequency components of a sequence, but also locate the timing of these frequency components [19]. By wavelet transform, the original highly variably emission time series/sequence can be reconstructed by the following wavelet coefficients as [20]

$$s[n] = \sum_j \sum_k c_{jk} \Psi(a_0^j nT - k) \tag{5}$$

where $s[n]$ is the original time series, $\{c_{jk}\}$ stands for a series of wavelet coefficients, $\Psi(a_0^j nT - k)$ is the wavelet at the j-th scale shifted by k samples, a_0 is a constant which is almost always set as 2. The original sequence can be reconstructed by different levels of high-frequency components D_i, and the residual low-frequency component A. The number of decomposition layers, j, must satisfy the wavelet variance test

$$\frac{std(A)}{std(s)} < 0.1, \tag{6}$$

where $std(\cdot)$ stands for the standard variation of a sequence.

Here we take the CO emission series of Lane 1 as an example. We use the 'db1' wavelet to decompose the original time series into four layers of high-frequency components D_i ($i = 1, 2, 3, 4$) and a layer of low-frequency component A, which are shown in Figure 3. We can compute $std(A)/std(s) = 0.03 < 0.1$, i.e., the requirement in (6) is satisfied. By processing D_i ($i = 1, 2, 3, 4$) and A with LSTM networks, we can forecast the accumulated hourly CO concentration as

$$CO[t] = \sum_{i=1}^{4} f(D_i[t-1], \ldots, D_i[t-5]) + f(A[t-1], \ldots, A[t-5]) \tag{7}$$

where $f(\cdot)$ is the LSTM network forecasting model based on lowly variable sub-series/sub-sequences. The details of $f(\cdot)$ will be presented in Section II-F.

F. LONG SHORT-TERM MEMORY NETWORK

The basic time series network model to predict emissions is shown in Figure 4, where 2 LSTM layers are connected to a fully connected neural network layer. In Figure 4, the pink part represents the input sequence, the green part is the LSTM layers, the yellow part is the fully connected neural network, and the blue part is the output. The dashed box indicates that the accumulated hourly concentrations can be iteratively predicted using the LSTM model.

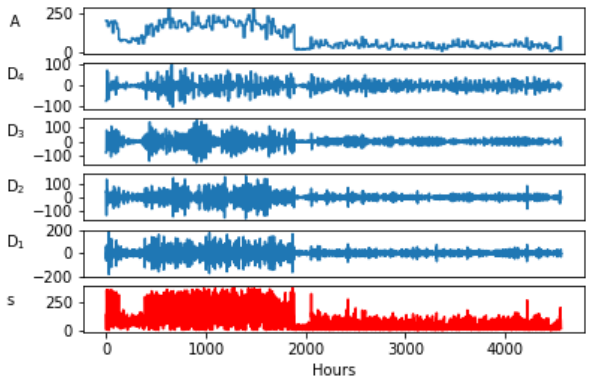

FIGURE 3. The wavelet decomposition of the original time series s of CO emission at Lane 1: D_1-D_4 denote the high-frequency components at different levels and A is the residual low-frequency component.

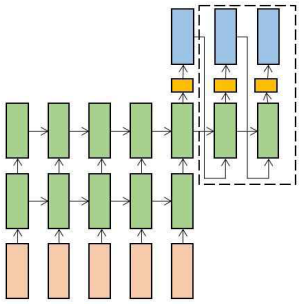

FIGURE 4. The basic structure of LSTM emission forecasting model.

The input of the model is a normalized 5×10 matrix, which are made up of the data in past 5 hours. More specifically, each row of that matrix is the hourly values of the number of vehicles passing by the monitoring lane, the average vehicle length, the average vehicle speed, the average vehicle acceleration, the wind speed, the wind direction, the outdoor temperature, the relative humidity, atmospheric pressure and the accumulated emission concentration of CO or HC or NO. Moreover, the input of that model is recurrent, i.e., the five consecutive inputs are sequentially got. The model output is the forecast accumulated hourly concentration of CO or HC or NO in the next hour (the dimension of the output is 1) or the next 3 hours (the dimension of the output is 3). LSTM is a special kind of recurrent neural network suitable for dealing with long-term dependence problems [21], [22]. Compared with the general recurrent neural networks, it adds a cell state that preserves long-term information in the hidden layer. The single layer LSTM structure is shown in Figure 5. After being temporally expanded, the single-layer LSTM structure is obtained and

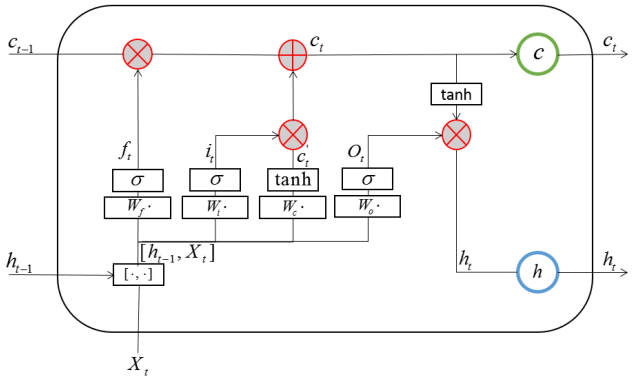

FIGURE 5. The single layer LSTM.

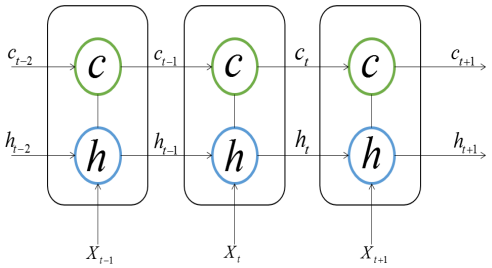

FIGURE 6. The temporal-logic diagram of a single layer LSTM.

shown in Figure 6.

In Figures 5 and 6, at time t, the LSTM input is the current time input vector X_t, the last time output vector is h_{t-1}, the last time cell state is c_{t-1}. In Figure 5, LSTM controls the cell state by the forget gate f_t and the input gate i_t. These gates are actually sigmoid functions that restrict the flow of information. The forget gate determines the information of the last cell state c_{t-1} to flow into the current cell state c_t and is expressed as

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f), \tag{8}$$

where $\sigma(\cdot)$ is the sigmoid nonlinear function, W_f is the weight matrix of the forget gate, b_f is the bias vector of the forget gate, $[h_{t-1}, X_t]$ is a vector to combine h_{t-1} and X_t.

The input gate i_t controls the information of the current input X_t to flow into the current cell state c_t, and is calculated as

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i), \tag{9}$$

where W_i and b_i are the weight matrix and the bias vector of the input gate, respectively.

The state of the current input, c'_t, can be calculated as

$$c'_t = tanh(W_c \cdot [h_{t-1}, X_t] + b_c), \tag{10}$$

where W_c is the weight matrix, b_c is the bias vector, and $tanh(\cdot)$ is the hyperbolic tangent function.

The current cell state c_t is determined by both the forget gate and the input gate as

$$c_t = f_t \circ c_{t-1} + i_t \circ c'_t, \tag{11}$$

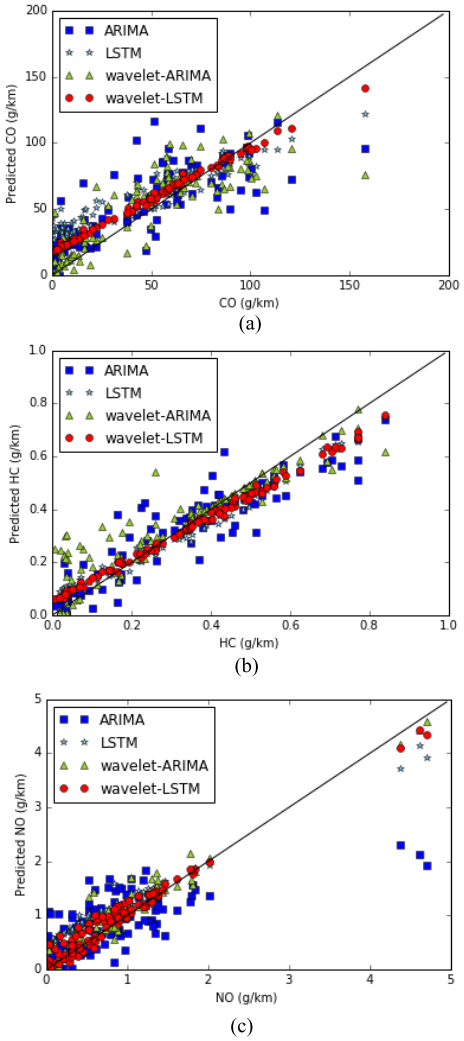

FIGURE 7. The 3-hour emission forecasting results for 100 consecutive samples from Lane 1. a) CO forecasting results. b) HC forecasting results. c) NO forecasting results.

where $\circ$ represents the element-wise multiplication between vectors.

The output gate O_t controls the information flowing from the cell state c_t, which contains the long-term memory information, to the current output, and is expressed as

$$O_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o), \tag{12}$$

where W_o and b_o are the weight matrix and bias vector, respectively.

The final output of the single LSTM layer is determined by the output gate O_t and the cell state c_t,

$$h_t = O_t \circ tanh(c_t) \tag{13}$$

The historical emission data are used to train the above wavelet-LSTM model. More specifically, the back propagation through time (BPTT) method is implemented to iteratively optimize the network weights, such as W_f, b_f, W_i, b_i, W_c, b_c, W_o, b_o, and the mean absolute error is selected as the network loss function.

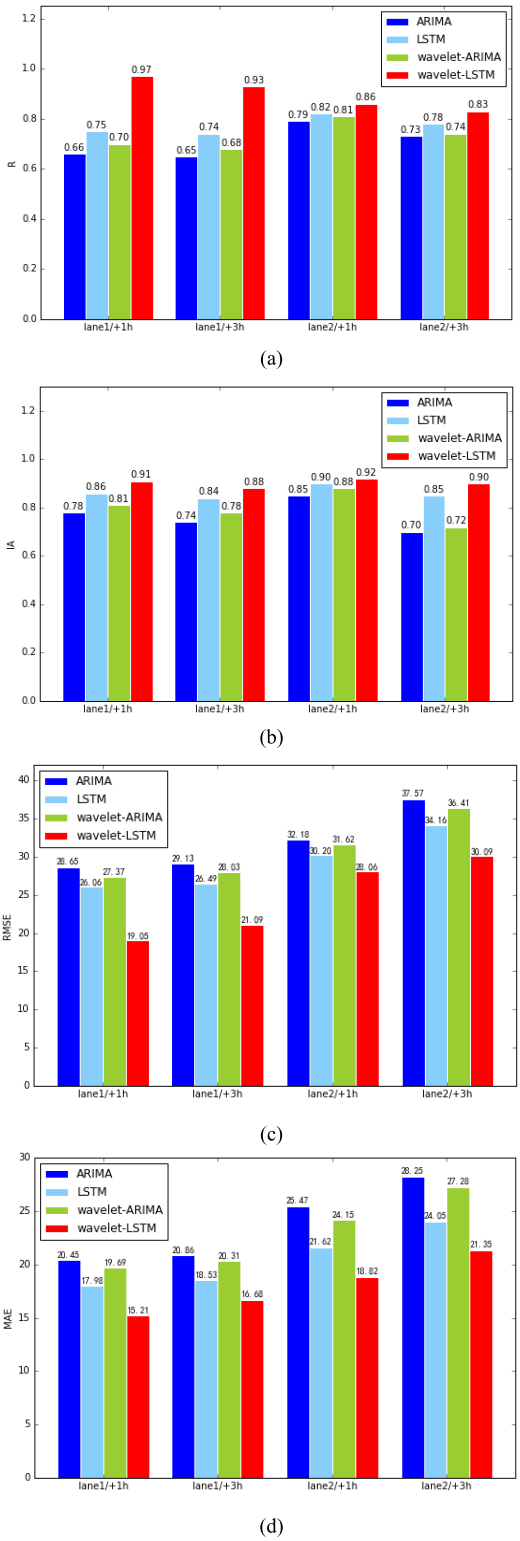

FIGURE 8. Performance of CO forecasting. a) *R* of CO forecasting. b) *IA* of CO forecasting. c) *RMSE* of CO forecasting. d) *MAE* of CO forecasting.

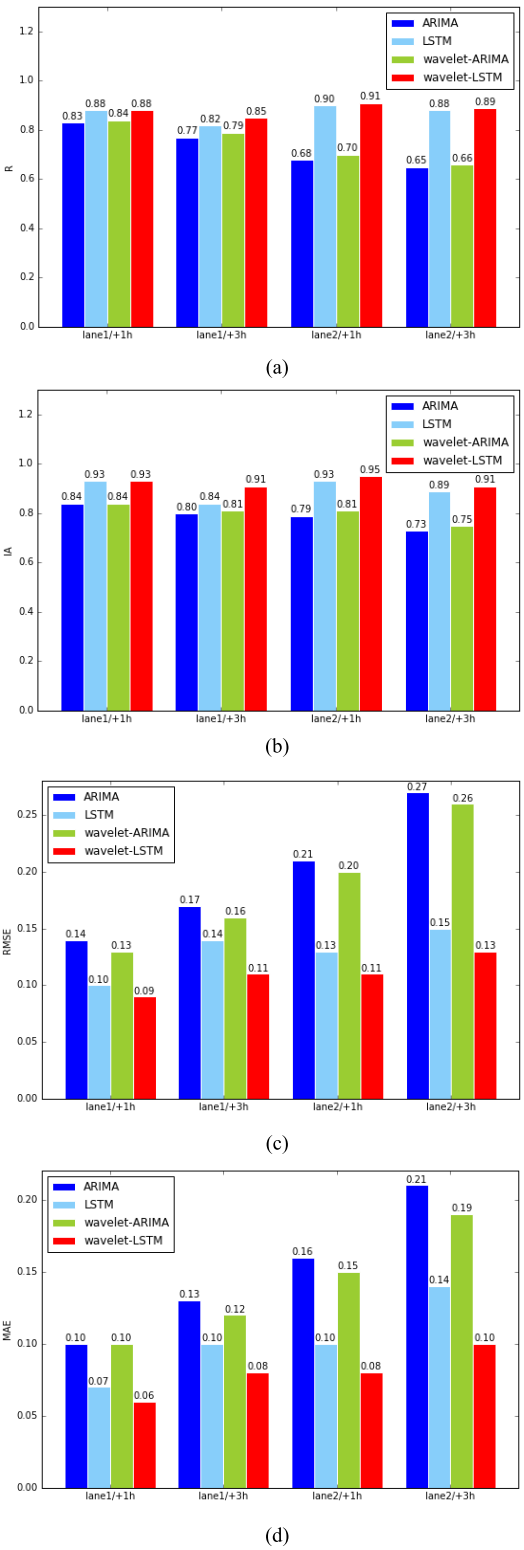

FIGURE 9. Performance of HC forecasting. a) *R* of HC forecasting. b) *IA* of HC forecasting. c) *RMSE* of HC forecasting. d) *MAE* of HC forecasting.

III. EXPERIMENTS AND DISCUSSION

Our emission forecasting model combines the wavelet transform in Section II-E and the LSTM network in Section II-F, and is referred to as wavelet-LSTM model. It is compared with 3 conventional forecasting models, including ARIMA, LSTM, wavelet-ARIMA which also introduces the wavelet

transform in Section II-E and implements the ARIMA model for each sub-series. As mentioned in Section II-A, the emission data at Lane 1 and Lane 2 were collected. With these emission data, the 4 forecasting models are trained and compared. Figure 7 shows the 3-hour emission forecasting results for 100 consecutive samples randomly selected from Lane 1. From Figure 7 we see that the predicted values are quite close to the real values under the proposed wavelet-LSTM model, which has achieved better performance than ARIMA, LSTM, and wavelet-ARIMA, especially during forecasting the concentration of CO in Figure 7-a).

Besides the qualitative comparison in Figure 7, the concerned 4 forecasting models are mainly quantitatively compared through a few indicators, including the correlation coefficient (R), the index of agreement (IA), the root mean square error (RMSE), and the mean absolute error (MAE), which are defined as

$$R = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2 \sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}, \tag{14}$$

$$IA = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(|y_i - \bar{y}| + |\hat{y}_i - \bar{y})^2}, \tag{15}$$

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

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|, \tag{17}$$

where N is the total number of samples in the test set, y_i is the real value of the i-th sample, $\bar{y}$ is the average of all samples, i.e., $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$, $\hat{y}_i$ is the forecasted value of the i-th sample, $\bar{\hat{y}}$ is the average of the forecasted values of all samples.

In our experiments, the 5-fold cross-validation method is implemented, i.e., the available emission data are separated into 5 equal sets in turn, 4 of which are used for training and the other one of which is used for validation or testing. The training and testing of each forecasting model is repeated for 5 times, each of which chooses a different set as the testing set. For each performance indicator, the average of the 5 obtained performance indicators is taken as the overall one. ARIMA, LSTM, wavelet-ARIMA, and wavelet-LSTM are implemented to forecast the accumulated hourly concentrations of CO, HC and NO in advance of one hour and three hours. The obtained performance results are shown in Figure 8, 9 and 10, where *"lane 1/+1h"* means the 1-hour forecasting at Lane 1 and *"lane 1/+3h"* means the 3-hour forecasting at Lane 1.

In Figure 8, the performance indicators of ARIMA, LSTM, wavelet-ARIMA and wavelet-LSTM are represented by blue, light blue, green and red bars, respectively. Figure 8-a) shows that R of the wavelet-LSTM model is largest and ARIMA obtains the smallest R. When forecasting the CO emission of Lane 1 in advance of 1 hour, R of the wavelet-LSTM model

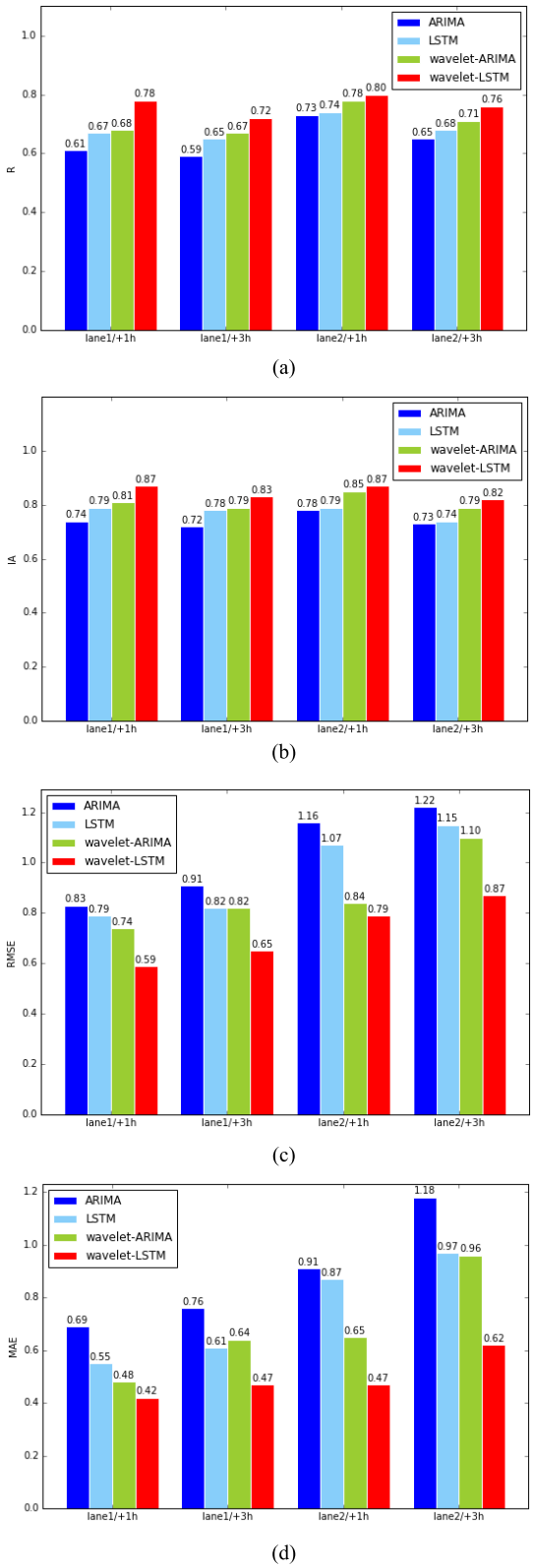

FIGURE 10. Performance of NO forecasting. a) *R* of NO forecasting. b) *IA* of NO forecasting. c) *RMSE* of NO forecasting. d) *MAE* of NO forecasting.

is as high as 0.97, which is much higher than 0.75 of the LSTM model, 0.70 of the wavelet-ARIMA model and 0.66 of the ARIMA model. Figure 8-b), 8-c) and 8-d) also confirm

TABLE 2. The forecasting results of Lane 1.

model			$\overline{y}$	R	IA	RMSE	MAE
CO	ARIMA	+1h	40.68	0.66	0.78	28.65	20.45
		+3h	40.68	0.65	0.74	29.13	20.86
	LSTM	+1h	40.68	0.75	0.86	26.06	17.98
		+3h	40.68	0.74	0.84	26.49	18.53
	wavelet-ARIMA	+1h	40.68	0.70	0.81	27.37	19.69
		+3h	40.68	0.68	0.78	28.03	20.31
	wavelet-LSTM	+1h	40.68	0.97	0.91	19.05	15.21
		+3h	40.68	0.93	0.88	21.09	16.68
HC	ARIMA	+1h	0.28	0.83	0.84	0.14	0.10
		+3h	0.28	0.77	0.80	0.17	0.13
	LSTM	+1h	0.28	0.88	0.93	0.10	0.07
		+3h	0.28	0.82	0.84	0.14	0.10
	wavelet-ARIMA	+1h	0.28	0.84	0.84	0.13	0.10
		+3h	0.28	0.79	0.81	0.16	0.12
	wavelet-LSTM	+1h	0.28	0.88	0.93	0.09	0.06
		+3h	0.28	0.85	0.91	0.11	0.08
NO	ARIMA	+1h	0.89	0.61	0.74	0.83	0.69
		+3h	0.89	0.59	0.72	0.91	0.76
	LSTM	+1h	0.89	0.67	0.79	0.79	0.55
		+3h	0.89	0.65	0.78	0.82	0.61
	wavelet-ARIMA	+1h	0.89	0.68	0.81	0.74	0.48
		+3h	0.89	0.67	0.79	0.82	0.64
	wavelet-LSTM	+1h	0.89	0.78	0.87	0.59	0.42
		+3h	0.89	0.72	0.83	0.65	0.47

TABLE 3. The forecasting results of Lane 2.

model			$\overline{y}$	R	IA	RMSE	MAE
CO	ARIMA	+1h	60.29	0.79	0.85	32.18	25.47
		+3h	60.29	0.73	0.70	37.57	28.25
	LSTM	+1h	60.29	0.82	0.90	30.20	21.62
		+3h	60.29	0.78	0.85	34.16	24.05
	wavelet-ARIMA	+1h	60.29	0.81	0.88	31.62	24.15
		+3h	60.29	0.74	0.82	36.41	27.28
	wavelet-LSTM	+1h	60.29	0.86	0.92	28.06	18.82
		+3h	60.29	0.83	0.90	30.09	21.35
HC	ARIMA	+1h	0.37	0.68	0.79	0.21	0.16
		+3h	0.37	0.65	0.73	0.27	0.21
	LSTM	+1h	0.37	0.90	0.93	0.13	0.10
		+3h	0.37	0.88	0.89	0.15	0.14
	wavelet-ARIMA	+1h	0.37	0.70	0.81	0.20	0.15
		+3h	0.37	0.66	0.75	0.26	0.19
	wavelet-LSTM	+1h	0.37	0.91	0.95	0.11	0.08
		+3h	0.37	0.89	0.91	0.13	0.10
NO	ARIMA	+1h	1.08	0.73	0.78	1.16	0.91
		+3h	1.08	0.65	0.73	1.22	1.18
	LSTM	+1h	1.08	0.74	0.79	1.07	0.87
		+3h	1.08	0.68	0.74	1.15	0.97
	wavelet-ARIMA	+1h	1.08	0.78	0.85	0.84	0.65
		+3h	1.08	0.71	0.79	1.10	0.96
	wavelet-LSTM	+1h	1.08	0.80	0.87	0.79	0.47
		+3h	1.08	0.76	0.82	0.87	0.62

the performance advantage of our wavelet-LSTM model in terms of *IA*, *RMSE* and *MAE*. The same performance ranking of ARIMA, LSTM, wavelet-ARIMA and wavelet-LSTM can also be observed in Figure 9 (HC forecasting), Figure 10 (NO forecasting). Moreover, we can observe from Figures 8-10 that

- *R* of 3-hour forecasting is smaller than that of 1-hour forecasting as forecasting becomes more difficult under a longer forecasting window.
- The wavelet transform does improve the performance by comparing ARIMA with wavelet-ARIMA, LSTM with wavelet-LSTM. This performance improvement mainly comes from the fact that the wavelet transform decomposes the highly variable original time series into lowly variable sub-series, and the forecasting based on these lowly variable sub-series is more efficient than the highly variable original time series, either by ARIMA or LSTM.
- LSTM achieves better forecasting performance than ARIMA by comparing LSTM with ARIMA, wavelet-LSTM with wavelet-ARIMA. The superiority of LSTM may come from its powerful modeling capability.

For precisely quantitative comparison, the performance results in Figure 8, 9, and 10 are listed in Tables 2 and 3. Table 2 shows the results of Lane 1 while Table 3 shows the results of Lane 2. As the concentration of CO is quite different from those of HC and NO, their averages of samples, $\bar{y}$, are included in Table 2 and 3. Besides the performance ranking of the 4 forecasting models, we observe from Tables 2 and 3 that Lane 2 suffers from worse pollution than Lane 1, i.e., higher concentrations of CO, HC and NO. This difference comes from the physical situation of the two lanes. Lane 2 lies in the right of Lane 1. According to the traffic rules in China, big vehicles, such as trucks, mainly drive on the most right lane, which is Lane 2 in our experiments. As big vehicles contribute most of CO, HC and NO, Lane 2 observes higher concentrations of CO, HC and NO.

IV. CONCLUSION

In this paper, a model based on wavelet transform and long short-term memory network is presented, which can accurately forecast the vehicle emissions of CO, HC and NO in advance. This model takes into account the number of vehicles passing by the monitoring lane, the average length of these vehicles, meteorological conditions and historical emissions. When some emission data samples are incomplete, a method based on semi-supervised collaborative training regression is used to compensate missing data. Wavelet transform is implemented to decompose the highly variable original time series of emission data into multiple lowly variable sub-series. For each sub-series, LSTM is implemented to forecast the concentrations of CO, HC and NO. By summing up the forecasted results of all sub-series, we obtain the overall concentrations of CO, HC and NO. Real emission data are used to train the proposed wavelet-LSTM model and compare it with conventional models, including ARIMA, LSTM and wavelet-ARIMA. Experimental results confirm that the proposed wavelet-LSTM model yields the best forecasting performance.

REFERENCES

[1] G. V. Franco, "Evaluation and improvement of road vehicle pollutant emission factors based on instantaneous emissions data processing," Ph.D. dissertation, Dept. Elect. Eng., Jaume I Univ., Castellón de la Plana, Spain, 2014.

[2] M. D. Geller, S. B. Sardar, H. Phuleria, and P. M. Fine, "Measurements of particle number and mass concentrations and size distributions in a tunnel environment," *Environ. Sci. Technol*, vol. 39, no. 22, pp. 8653–8663, 2005.

[3] J. Merkisz, J. Pielecha, P. Fuc, and P. Lijewski, "The analysis of the PEMS measurements of the exhaust emissions from city buses using different research procedures," in *Proc. IEEE VPPC*, Seoul, South Korea, Oct. 2012, pp. 903–907.

[4] Z. Ning and T. L. Chan, "On-road remote sensing of liquefied petroleum gas (LPG) vehicle emissions measurement and emission factors estimation," *Atmos. Environ.*, vol. 41, no. 39, pp. 9099–9110, 2007.

[5] M. Shaharuddin, A. Zaharim, M. J. M. Nor, O. A. Karim, and K. Sopian, "Application of wavelet transform on airborne suspended particulate matter and meteorological temporal variations," *WSEAS Trans. Environ. Develop.*, vol. 4, no. 2, pp. 89–98, 2008.

[6] X. H. Cai, "Time series analysis of air pollution co in California south coast area, with seasonal ARIMA model and VAR model," Ph.D. dissertation, Dept. Elect. Eng., Univ. California, Los Angeles, CA, USA, 2009.

[7] J. Kukkonen *et al.*, "Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentrations, compared with a deterministic modelling system and measurements in central Helsinki," *Atmos. Environ.*, vol. 37, no. 32, pp. 4539–4550, 2003.

[8] M. Cai, Y. Yin, and M. Xie, "Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach," *Transp. Res. D, Transport Environ.*, vol. 14, no. 1, pp. 32–41, 2009.

[9] J. Connor and L. Atlas, "Recurrent neural networks and time series prediction," in *Proc. Seattle Int. Joint Conf. Neural Netw.*, vol. 1, Jul. 1991, pp. 301–306.

[10] U. Brunelli, V. Piazza, L. Pignato, L. Pignato, and S. Vitabile, "Two-days ahead prediction of daily maximum concentrations of SO_2, O_3, PM_{10}, NO_2, CO in the urban area of Palermo, Italy," *Atmos. Environ.*, vol. 41, no. 14, pp. 2967–2995, 2007.

[11] M. J. Alizadeh, E. J. Nodoushan, N. Kalarestaghi, and W. Chau, "Toward multi-day-ahead forecasting of suspended sediment concentration using ensemble models," *Environ. Sci. Pollut. Res.*, vol. 24, no. 36, pp. 28017–28025, 2017.

[12] M. J. Alizadeh, M. RezaKavianpour, O. Kisi, and V. Nourani, "A new approach for simulating and forecasting the rainfall-runoff process within the next two months," *J. Hydrol.*, vol. 548, pp. 588–597, May 2017.

[13] L. Salazar, O. Nicolis, F. Ruggeri, J. Kisel'ák, and M. Stehlík "Predicting hourly ozone concentrations using wavelets and ARIMA models," *Neural Comput. Appl.*, pp. 1–10, Jan. 2018.

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[15] Y.-W. Ko and C.-H. Cho, "Characterization of large fleets of vehicle exhaust emissions in middle Taiwan by remote sensing," *Sci. Total. Environ.*, vol. 354, no. 1, pp. 75–82, 2006.

[16] B. A. Holmén and D. A. Niemeier, "Characterizing the effects of driver variability on real-world vehicle emissions," *Transp. Res. D, Transport Environ.*, vol. 3, no. 2, pp. 117–128, 1998.

[17] Z. H. Zhou and M. Li, "Semisupervised regression with cotraining-style algorithms," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 11, pp. 1479–1493, Nov. 2007.

[18] T. L. Chan, Z. Ning, C. W. Leung, C. S. Cheung, W. T. Hung, and G. Dong, "On-road remote sensing of petrol vehicle emissions measurement and emission factors estimation in Hong Kong," *Atmos. Environ.*, vol. 38, no. 14, pp. 2055–2066, 2004.

[19] T. Partal, "River flow forecasting using different artificial neural network algorithms and wavelet transform," *Can. J. Civil. Eng.*, vol. 36, no. 1, pp. 26–38, 2009.

[20] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61. Philadelphia, PA, USA: SIAM, 1992.

[21] H. Sak, A. Senior, and F. Beaufays "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *Proc. 15th Annu. Conf. Int. Speech Commun. Assoc.*, Singapore, 2014, pp. 338–342.
[22] K. Kawakami, "Supervised sequence labelling with recurrent neural networks," Ph.D. dissertation, Dept. Elect. Eng., Tech. Univ. Munich, Munich, Germany, 2008.

QIANG ZHANG received the B.E. degree from the University of Science and Technology of China, Hefei, China, in 2016, where he is currently pursuing the M.S. degree with the Department of Automation. His research interests include signal processing and machine learning.

FENG LI received the B.S., M.Eng., and Ph.D. degrees from the University of Science and Technology of China, Hefei, China, in 2000, 2006, and 2011, respectively. He is currently a Lecturer with the University of Science and Technology of China. His research interests include signal processing and intelligent video surveillance systems.

FEI LONG received the B.Sc. degree from Nanjing University in 2005 and the Ph.D. degree from Tsinghua University in 2010. He was a Visiting Scholar with the Computer Science department, Yale University, from 2007 to 2009. He was a Post-Doctoral Fellow with The Hong Kong University of Science and Technology. He joined the 54th Research Institute of China Electronics Technology Group Corporation in 2011, and was promoted as a Senior Engineer in 2012. He was a Chief Designer of a series of important projects about delay tolerant networks and satellite networks. He joined Chinaso, Inc., in 2013, as the Director of the Innovative Development Department, and was in charge of the R&D of frontier technology of Internet. He organized and accomplished the scholar, book, and shitu channels of Chinaso, Inc. He has published over 20 papers, three monographs, and two translations. His research interests span a wide range of topics in network routing, deep learning, and pattern recognition.

QIANG LING (M'05–SM'10) received the B.S. degree from the University of Science and Technology of China, Hefei, China, in 1997, the M.E. degree from Tsinghua University, Beijing, China, in 2000, and the Ph.D. degree from the University of Notre Dame, Notre Dame, IN, USA, in 2005. He was a Research Staff Member with Seagate Technology from 2005 to 2008. He joined the University of Science and Technology of China in 2008, where he is currently an Associate Professor with the Department of Automation. His research interests include networked control systems and signal processing. He is currently serving as an Associate Editor on the IEEE Control Systems Society Conference Editorial Board.