

# Near-road fine particulate matter concentration estimation using artificial neural network approach

D. Z. Zhang · Z. R. Peng

Received: 8 January 2014 / Revised: 25 February 2014 / Accepted: 17 March 2014 / Published online: 29 April 2014  
© Islamic Azad University (IAU) 2014

**Abstract** Evidence has shown a strong association between ambient particulate matter and adverse health problems. In urban areas, most of households are located near arterial roads, which are exposure to fine particulate matter directly. Hence, it is critical to understand the near-road fine particulate matter concentration and distribution for the purpose of health risk analysis. This paper applies artificial neural network to estimate the near-road fine particulate matter concentration. Factors influencing the detected concentration are classified into four categories: traffic-related, weather-related, detection location-related and background-related. The estimated values are compared with concentrations detected by monitoring campaigns in Gainesville, FL and Shanghai, China. Distinguished from previous research, this study illustrates the fine particulate matter dispersion and distribution within 50 m near road with portable fine particulate matter detectors and weather instruments. The results indicate that artificial neural network approach is capable of producing accurate estimation of pollutant dispersion near road. Besides, fine particulate matter concentration decayed about a half at 30 m distance from an arterial road in Gainesville, FL. Background contributes to more than 2/3 of the detected value at roadside in Shanghai, and the distance–decay pattern is not as obvious as that in

Gainesville, which is different from previous studies reported in the literature. An artificial neural network model performs better after removing the background concentration and with higher concentration value of fine particulate matter.

**Keywords** Fine particulate matter · Artificial neural network · Dispersion prediction model · Monitoring campaign

## Introduction

Vehicular exhaust emission is one of the major sources of air pollution within urban area. The combustion processes in diesel and gasoline engines contributes significantly to particulate matter (PM), especially to particles with diameters smaller than  $2.5\ \mu\text{m}$  (fine particulate matter,  $\text{PM}_{2.5}$ ) (Lighty et al. 2000; US Department of Energy 2010). It is reported that the health problems associated with  $\text{PM}_{2.5}$  and other pollutions accounts for 2 % of cardiorespiratory mortality and 5 % of respiratory cancers (World Health Organization 2002). The connection between  $\text{PM}_{2.5}$  exposure and increasing adverse health problem is under increasing scrutiny of the general public (Abdullah et al. 2007; Zhao et al. 2013). It is reported that approximately 11 % of US households are located within 100 m of four-lane highways; more than half of living neighborhoods locates within the 3 km of urban arterials in China (Brugge et al. 2007; Cai et al. 2009). Hence, using appropriate models with vehicle and meteorological information to understand near-road  $\text{PM}_{2.5}$  concentration and its dispersion is critical and essential to neighbors nearby.

Cai et al. (2009) indicated that the dispersion characteristics of  $\text{PM}_{2.5}$  near road are highly nonlinear and are

---

D. Z. Zhang · Z. R. Peng  
Center for ITS and UAV Applications Research, State Key Laboratory of Ocean Engineering, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China

Z. R. Peng (✉)  
Department of Urban and Regional Planning, University of Florida, Gainesville, FL 32611-5706, USA  
e-mail: zrpeng@sjtu.edu.cn



closely related with traffic emission, background concentration, meteorological and geographical conditions, and some other local characteristics. Due to its complexity, widely studied models, such as CALINE4, CAL3QHC and AERMOD showed limitation when applying for near-road prediction (Chen et al. 2009; Yura et al. 2007). On the other hand, artificial neural network (ANN) is particularly suitable for modeling multifactor, uncertainty and nonlinearity (Kukkonen et al. 2003). Hence, ANN-based models are widely used to estimate air pollutant concentrations such as NO<sub>2</sub>, PM<sub>10</sub>, SO<sub>2</sub>, ozone and PM<sub>2.5</sub> (Bandyopadhyay and Chattopadhyay 2007; Nagendra and Khare 2006; Ordieres et al. 2005).

In this research, with the portable PM detector and weather station, the local weather information and PM<sub>2.5</sub> dispersion concentrations in various detecting points have been obtained, which provides better observation data comparing with previous research reported in the literature. Meanwhile, monitoring campaigns were run in US and China to compare the estimation performance of different ANN model structures.

This paper presents the results of on-site monitoring campaigns of near-road PM<sub>2.5</sub> concentrations in Gainesville, United States and Shanghai, China. The monitoring campaigns run from January to July, 2013, which is detailed in materials and methods section. Then, the ANN-based models are established and analyzed for near-road PM<sub>2.5</sub> estimation in results and discussion section. Finally, in the last section, conclusion and limitations of this research are discussed.

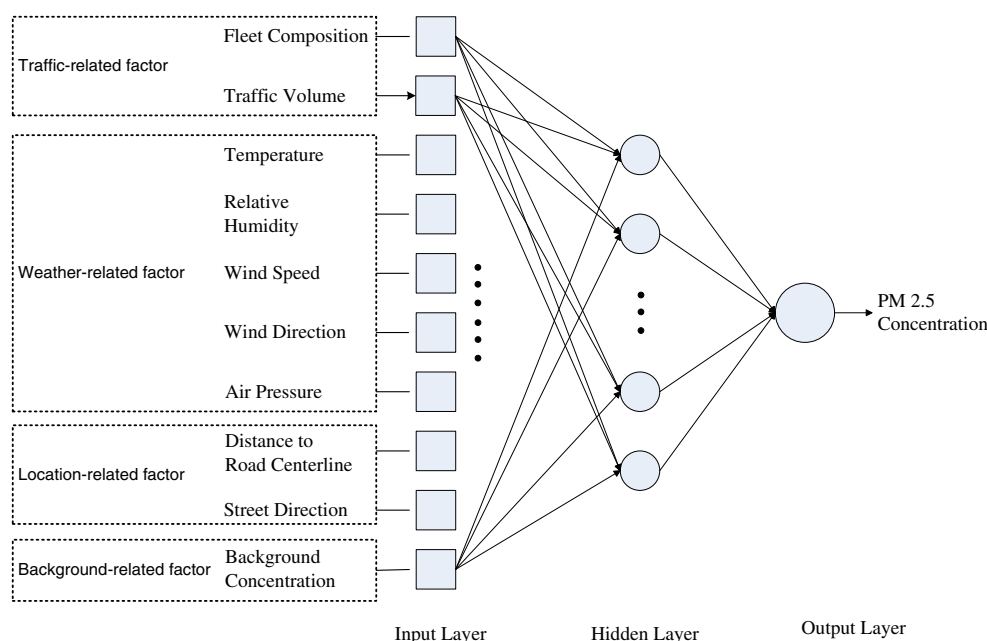
## Materials and methods

### Methodology

Artificial neural networks are computational models based on biological neural network. Consisting of mutual connected neurons hidden in the system, they process the information through a step-learning phase. Different from other modeling methods, ANNs make no prior assumptions about the variable distribution; besides, the inherent propensity for storing empirical knowledge and abilities of learning and generalizing about massive amount of data make ANNs capable of modeling highly nonlinear relationship (Gardner and Dorling 1998; Karlaftis and Vlahogianni 2011). There are several kinds of ANN-based models, among which back-propagation (BP) neural network is one of the powerful and most widely used ones (Rumelhart et al. 1986). BP algorithm is a supervised learning approach by computing the error gradient for a feed-forward network. Feed-forward networks own one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons, which are desirable for solving nonlinear problem. Figure 1 showed the ANN-based neural network structure used in this paper. Input layer, hidden layer and output layer were connected by synaptic weights.

In this paper, a widely-recognized sigmoid transfer function was adopted:

$$y_j = f(\text{net}_j) = \frac{1}{1 + e^{-\text{net}_j}} \quad (1)$$



**Fig. 1** Architecture of the proposed ANN-based prediction model



where,  $net_j$  was the state of the  $j$ -th neuron in the hidden layer, and

$$net_j = \sum_i w_{ji}x_i - \theta_j \quad (2)$$

where,  $w_{ji}$  was the weight of the  $i$ -th neuron of the input layer to the  $j$ -th neuron in the hidden layer;  $x_i$  is the output of the  $i$ -th neuron of the input layer;  $\theta_j$  is the bias invariant of the  $j$ -th neuron of the hidden layer. Consequently, the input and output of the hidden layer and the output layer could be obtained:

$$\begin{aligned} hi_h(k) &= \sum_{i=1}^n w_{ih}x_i(k) - b_h, ho_h(k) = f(hi_h(k)) \\ &= 1/(1 + \exp(-\sum_{i=1}^n w_{ih}x_i(k) - b_h)) \end{aligned} \quad (3)$$

$$\begin{aligned} yi_o(k) &= \sum_{h=1}^p w_{ho}ho_h(k) - b_o, yo_o(k) = f(yi_o(k)) \\ &= 1/(1 + \exp(-\sum_{h=1}^p w_{ho}ho_h(k) - b_o)) \end{aligned} \quad (4)$$

where,  $hi_h(k)$  and  $ho_h(k)$  were the input and output of the hidden layer;  $yi_o(k)$  and  $yo_o(k)$  were the input and output of the output layer.

The total error of the neural network could be expressed as follows:

$$E = \sum_k E^k, E^k = 1/2[d_o(k) - yo_o(k)]^2 \quad (5)$$

where,  $d_o(k)$  was the desired value of the  $k$ -th sample.

According to the total error, the modified increments  $\Delta w_{ho}(k)$  of the weight between  $k$ -th hidden neuron and the output neuron could be determined:

$$\Delta w_{ho}(k) = -\mu \frac{\partial E^k}{\partial w_{ho}} = \mu \delta_o(k) ho_h(k) \quad (6)$$

where,  $\mu$  was the learning rate;  $\delta_o(k)$  was the partial derivative of the error  $E^k$  to  $yi_o(k)$ . The modified weight could be determined as  $w_{ho}^{N+1} = w_{ho}^N + \varphi \delta_o(k) ho_h(k)$ .

Similarly, the modified increment  $\Delta w_{ih}(k)$  of the weight between the  $i$ -th input neuron and the  $k$ -th hidden neuron could be determined:

$$\Delta w_{ih}(k) = -\mu \frac{\partial E^k}{\partial w_{ih}} = -\mu \frac{\partial E^k}{\partial hi_h(k)} \frac{\partial hi_h(k)}{\partial w_{ih}} = \delta_h(k) x_i(k) \quad (7)$$

and the modified weight  $w_{ih}^{N+1} = w_{ih}^N + \varphi \delta_h(k) x_i(k)$ .

In Eqs. (6) and (7), learning rate  $\mu$  and  $\varphi$  controlled the convergence speed to the minimum of errors. In this paper,  $\mu = \varphi = 0.3$  were adopted in the BP neural network based on previous research results (Cai et al. 2009).

## Influential factors selection

Factors that influenced the dispersion and concentrations of air pollutants near urban arterials could be classified into four aspects: traffic-related, weather-related, location-related and background-related. Based on the classification, ten factors were selected into the input layer of ANN-based prediction model, as shown in Fig. 1.

### Traffic-related factors

Vehicular emission is one of the major reasons that contribute to near-road  $PM_{2.5}$  pollution. Previous studies indicated that traffic volume, speed, vehicle type and other factors have a significant impact on near-road  $PM_{2.5}$  concentration (Kam et al. 2012). In this research, factors such as traffic volume and vehicle type were considered based on on-site video recording. Fleet composition was obtained by calculating the ratio of heavy-duty vehicles to total number of vehicles including both light-duty and heavy-duty vehicles. The vehicle types used in this research were defined by the governments of both United States and China (National Research Council 2000; Ministry of Housing and Urban–Rural Construction of the P.R. China 2012). The vehicle speed was not taken into consideration because the average vehicle speed profiles across times of the day were pretty stable in both roads selected for monitoring. Therefore, traffic-related factors included in this research contained traffic volume and fleet composition.

### Weather-related factors

Meteorology factors could directly influence the air pollutant dispersion. Factors such as temperature, relative humidity, wind speed, wind direction, air pressure were usually used in the previous research about pollutant dispersion. With the portable weather station, micro-level meteorology data could be obtained, which could reflect the local weather condition during the monitoring campaigns.

### Location-related factors

The geographical location of monitoring site was another factor for pollutant concentration. Air pollutant caused by transportation near urban arterials belongs to line source dispersion problem. Hence, the relation between wind direction and road coordinate system is important for predicting  $PM_{2.5}$  concentration. Besides, there was a distance–decay gradient in  $PM_{2.5}$  concentration near urban arterials (Beckerman et al. 2008). Therefore, factors of distance to





**Fig. 2** Study areas and monitoring sites in **a** Gainesville, FL and **b** Shanghai, China

road centerline and street direction were included in location-related factors for the model.

#### Background-related factors

Previous researches indicated that background concentration accounted for the majority of the observed concentration. Hence, removing the background concentration was one effective approach to analyze other influential factors' impact on air pollution (Cai et al. 2009). In this research, instead of traditional fixed station detection, a micro-level background concentration could be obtained and used as input variable based on portable  $PM_{2.5}$  measures.

#### Case study: field experiments

##### Study areas

Data collected from two monitoring sites were used to evaluate the capabilities of the ANN-based prediction model in this research. One site is near the SW Archer Rd. in Gainesville, Florida (Fig. 2a); the other is near the Jianchuan Rd. in Shanghai (Fig. 2b). Sites were selected based on the following criteria: (1) road with relative high traffic volume; (2) open road without buildings and barriers nearby; and (3) places without pollutant influence from surroundings, such as park lots and industrial plants.

SW Archer Rd. is a four-lane road (both direction) with a total width of 25 m. The average annual daily traffic volume on this road was 19,200 passenger car units (PCUs) based on 2012 Traffic Report of Alachua County (Florida DOT 2013). As shown in Fig. 2a, angle  $\alpha$  was  $45^\circ$ , and six

detecting points were set in perpendicular direction along the road in order to monitor the distance decay without significant influence from surrounding roadways or stationary emission sources.

Jianchuan Rd. is an arterial in the southwestern Shanghai with four lanes in total. The total width is 30 m with an average traffic volume of approximately 1,500 vehicles per hour in the daytime during the monitoring campaigns. Six detecting points were selected away from the arterial, as shown in Fig. 2b, angle was  $30^\circ$ .

For both monitoring campaigns, one portable  $PM_{2.5}$  detector was used for measuring the  $PM_{2.5}$  concentration along perpendicular direction with the distance to road centerline of: 15, 20, 25, 30, 40 and 50 m. A fixed interval measurement was conducted for each detecting point. The local background concentration was monitored by other two detectors located more than 100 m away from the arterials, avoiding impacts of surroundings such as parking lots, industrial plants et al. The average value of the two detectors was adopted as the local background concentration.

#### Data collection

The monitoring campaigns ran from January to July, 2013. For both sites, data were collected in three time periods, morning (9.00–12.00 a.m.), afternoon (1.00–4.00 p.m.) and evening (5.00–7.00 p.m.)

The  $PM_{2.5}$  concentrations were measured using TSI Sidepak AM510 Detector, a portable device designed to monitor and record fine PM concentrations on a second-by-second basis (Fig. 3a). With light-scattering techniques, it is capable of measuring  $PM_{2.5}$  concentrations between 1





**Fig. 3** Monitoring instruments:  
**a** Portable PM<sub>2.5</sub> detector;  
**b** Weather station; **c** On-site monitoring campaign



(a)



(b)



(c)

**Table 1** Details of the monitoring instruments

Parameter	Range	Resolution	Accuracy	Unit
PM <sub>2.5</sub>	1–20,000	1	±1 µg/m <sup>3</sup>	µg/m <sup>3</sup>
Wind speed	1–67	0.5 m/s	±5 %	m/s
Wind direction	0–360	1°	±3°	Degree
Temperature	–40–65	0.1	±0.5 °C	°C
Relative humidity(RH)	0–100	1 %	±3–4 %	%
Barometric pressure	540–1,100	0.1 hPa	±1.0 hPa	hPa

and 20,000 µg/m<sup>3</sup> with an accuracy of 1 µg/m<sup>3</sup>. Weather-related parameters, such as temperature, wind speed, wind direction, air pressure et al., were measured and collected by Davis Vantage VUE Weather Station, as shown in Fig. 3b. Details of the instruments for each parameter were listed in Table 1. All sensors used in the experiments were factory calibrated within 6 months prior to the start of the experiments.

The traffic condition of both sites were recorded by video and counted by volume counting software afterward. One-minute traffic volumes during the monitoring period were counted by light-duty vehicles and heavy-duty vehicles, respectively. The fleet composition was reflected by the percentage of heavy-duty vehicles.

On-site monitoring campaigns lasted for more than 5 months. Observed concentrations from 21 days were collected and used for model estimation, including 15 weekdays and 6 weekends. Table 2 presented the background and increment concentrations during the monitoring campaign. The monitoring in US was conducted in January and February in Gainesville, FL. The average temperature during this period was 16 °C, ranging from 5 to 23 °C. The other part of experiments was taken place in

Shanghai from March to July, 2013, with an average temperature of 20 °C, ranging from 4 to 33 °C.

### Exploratory analyses

During the monitoring campaign, more than 1,200 records were obtained, including traffic volume, PM concentration, weather conditions and background concentration. The data collected in Gainesville indicated that 82.2 % of total samples in US had the value less than 20 µg/m<sup>3</sup>, while the data in Shanghai showed that 93.6 % of the total sample had the concentration higher than 50 µg/m<sup>3</sup>. The average concentration of all samples collected in Gainesville is 16.2 µg/m<sup>3</sup>, while in Shanghai, the average value is 108.6 µg/m<sup>3</sup>. General conclusion from the recorded sample is that regardless of influence from all other factors, the overall air quality in Gainesville is more than six times better than that of Shanghai.

Additionally, the percentages of background concentration to observed concentration are various for both sites. Background concentration accounts for 28.9 % of the average detected concentration in Gainesville during the monitoring days, while, in Shanghai, it accounts for 70.8 % of average concentration of the total detected.

Figure 4 showed the observed PM<sub>2.5</sub> concentration near road. The concentration decay from the road centerline was obvious in Gainesville's case. Such findings are consistent with previous research (Beckerman et al. 2008), indicating a distance–decay in pollutant dispersion near expressways. The detector in 30 m away from the road centerline showed a 50 % decay in PM<sub>2.5</sub> concentration. On the contrary, concentration data in Shanghai show rather weak relationship of distance–decay. The reason came from relative high background concentration in Shanghai, which reduced the pollution impact from the



**Table 2** Background and increment PM<sub>2.5</sub> concentrations in monitoring

Location	Time	Background (μg/m <sup>3</sup> )	Increment (μg/m <sup>3</sup> )					
			15 m	20 m	25 m	30 m	40 m	50 m
SW Archer Rd., Gainesville, United States	Feb. 12th, 09:00–10:00	8.7	17.07	14.91	10.05	9.30	5.48	3.27
	Feb. 15th, 10:00–11:00	2.3	12.31	9.82	9.65	8.23	5.60	3.84
	Feb. 17th, 12:45–13:45	1.36	8.96	5.99	5.47	4.04	2.82	1.64
	Feb. 18th, 14:05–15:05	1.15	8.63	5.81	5.59	3.98	1.16	−0.24 <sup>a</sup>
	Feb. 20th, 17:00–18:00	10.3	− <sup>b</sup>	72.18	34.02	5.91	−2.21 <sup>a</sup>	0.79
	Feb. 21st, 17:00–18:00	9.2	− <sup>b</sup>	63.89	36.83	4.52	2.10	2.12
Jianchuan Rd., Shanghai, China	Mar. 28th, 12:15–13:15	53.7	24.3	21.9	23.0	24.5	25.8	27.7
	Mar. 29th, 13:40–14:40	62.1	14.2	9.8	13.4	11.7	14.1	9.5
	Mar. 30th, 14:45–15:45	55.3	18.3	15.7	11.6	11.0	10.3	7.0
	Apr. 1st, 15:50–16:50	47.8	15.0	12.0	9.3	8.1	5.0	5.2
	Apr. 3rd, 16:45–17:45	39.2	13.0	18.5	11.1	8.3	7.7	4.7
	Apr. 20th, 9:05–10:05	78.1	57.1	61.0	54.6	64.8	58.1	58.4
	Apr. 21st, 10:00–11:00	83.1	57.4	50.9	58.7	64.7	68.8	65.8
	Apr. 22nd, 11:00–12:00	96.4	62.0	65.2	75.3	82.7	83.1	84.7
	Apr. 27th, 12:00–13:00	125.8	82.0	92.1	102.8	98.0	98.6	94.7
	Apr. 30th, 13:00–14:00	158.2	68.9	62.0	69.0	81.1	95.6	97.0
	May. 12th, 12:45–13:45	68.1	21.5	12.1	16.0	17.4	26.4	25.9
	Jun. 8th, 13:45–14:45	81.2	24.8	17.4	13.7	10.6	8.7	5.2
	Jun. 10th, 14:45–15:15	76.5	18.9	15.0	79.3	1.0	118.0	119.0
	Jul. 14th, 12:15–15:15	21.2	10.4	9.2	9.1	9.1	8.5	8.4
	Jul. 16th, 17:50–20:50	34.9	10.8	10.1	7.3	8.5	7.2	6.5

All data were collected minute by minute. The average concentrations during monitoring period were calculated for each detection point

<sup>a</sup> The increment were negative when the detected concentration minus the background. It indicated that the detected value was nearly the same as background value or even lower than the background. In such case, local weather or other unpredicted factor might contribute to the detected concentration

<sup>b</sup> Data were missing due to barrier's occupying the desirable detection place

traffic volume. Turbulence caused by wind was another reason leading to the slightly lower concentrations observed near road.

The pattern in Fig. 4 was consistent with past studies (Lowenthal et al. 2013), showing average PM<sub>2.5</sub> concentrations in the morning were found to be significantly higher than those in the afternoon or evening. In Gainesville, the monitor campaigns were held in peak hour (17:00–19:00) with relative high traffic volume compared with monitoring in the morning or afternoon (the average traffic volume is 3,388 vehicles per hour, 40 % higher than that of other monitoring periods). The monitoring campaign's data in Shanghai indicated that traffic volume in the morning is 1,833 vehicles per hour, which is 28 and 15 % higher than those in the afternoon and in the evening. Conclusion could be obtained that traffic-related influence is one of the dominant factors leading to high observed PM<sub>2.5</sub> concentrations within 15 m to 30 m near road when ignoring the background concentration's influence.

## Results and discussion

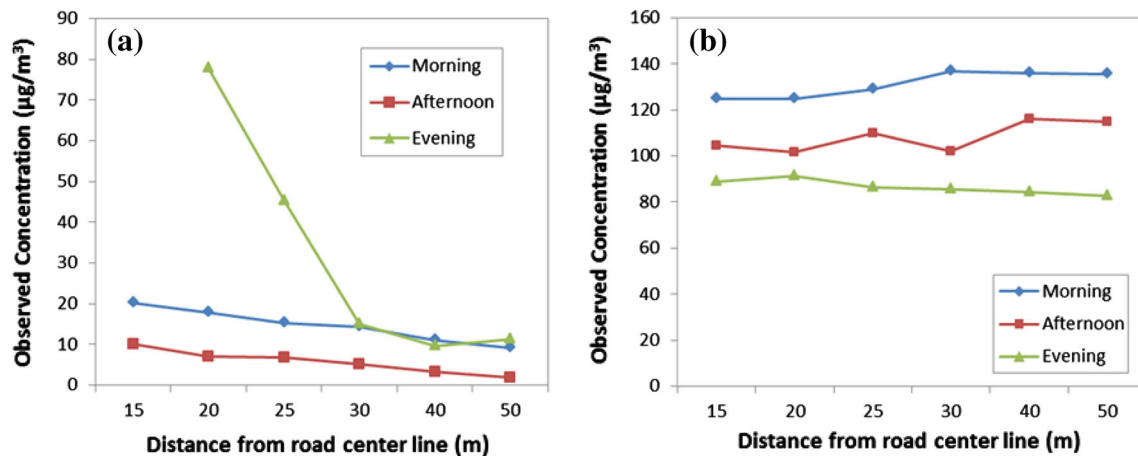
### Sample classification and normalization

The valid dataset contained 973 samples, which were further divided into three subsets, namely training set (75 % of the samples), evaluation set (10 % of the samples) and testing set (15 % of the samples). The training dataset used for model self-learning process, then adjusted according to the error; the evaluation dataset used to measure network generalization and halted the training when model performance stopping improving; testing set used to assess the performance of the network after learning.

To avoid overflows of neural network because of extreme weights and eliminate the influence of different dimensions of data, the input variables of neural network needed to be normalized (Gardner and Dorling 1998). In this paper, the input variables were normalized as defined in Eq. (8).

$$X_{\text{norm}}^i = (X_i^k - X_{\text{min}}^k) / (X_{\text{max}}^k - X_{\text{min}}^k) \quad (8)$$





**Fig. 4** Near-road observed concentrations: **a** Archer Rd., Gainesville; **b** Jianchuan Rd., Shanghai

where,  $X_{\text{norm}}^i$  was the normalized value,  $X_i^k$  was the original input value of factor  $k$ ,  $X_{\text{min}}^k$  and  $X_{\text{max}}^k$  were the minimum and maximum values of factor  $k$ .

#### Building neural network

When building the feed-forward neural network, number of hidden layers was one of significant factors. Research indicated that a network with single hidden layer including sufficiently large number of neurons could approximate any measurable and smooth function between input and output factors (Hornik et al. 1989). Hence, a single layer neural network was introduced into this paper.

Neural network was sensitive to the number of neurons in the hidden layers. Too few neurons could lead to under-fitting, while, too many could contribute to oscillation of fitting curve, thereby reducing the generalization capacity of a network (Hagan et al. 1996). Considering the surrounding diversity, in this paper, the number of hidden neurons was determined respectively for both monitoring sites. Mean relative error (MRE), mean absolute error (MAE) and mean square error (MSE) were used to evaluate the estimation results, which were shown in Table 3. It could be found that with ten hidden neurons, the neural network produces the best estimation in Shanghai. Meanwhile, best performance is observed with 12 hidden neurons in Gainesville, FL (as shown in bold in Table 3).

#### Influence of background concentration on the ANN-based model

From the exploratory analysis above, the data indicated that background concentration accounted for more than 2/3 of detected concentration in Shanghai. Hence, it was critical to check the influence of background concentration to the prediction model.

Factor-of-two plot is a classical method to reflect model performance. Typically, if 80 % of the points fall inside the factor-of-two envelope, the model results are considered good in estimating true value (Yura et al. 2007). Figure 5 presented that, in Gainesville, 91.7 % of total points fell inside the factor-of-two envelope when considering the background influence, while 85.9 % were inside the envelope if kicking off the influence of background concentration. The percentages falling inside the envelopes in Shanghai case were 100 and 93.0 %, respectively. Results showed that considering the background, the percentages of inside points were higher for both cases. However, with the limitation of factor-of-two analysis (for low increment concentrations, a point with good estimation may lie outside the envelope), a conclusive argument for model effectiveness could not be obtained (Cai et al. 2009).

The performance of ANN-based model could be assessed numerically by descriptive statistics. Table 4 details the statistics measure of ANN-based model estimation compared with observed concentrations. Except for ME, MAE, RMSE mentioned above, measures as  $R^2$ ,  $T$  and Durbin-Watson ( $S_{D-W}$ ) statistic were taken into consideration.  $R^2$  is one of the major measures indicating the performance of prediction model; Theil's inequality coefficient ( $T$ ), as shown in Eq. (9), is a measure of a time series of estimated values compares to observed values;  $S_{D-W}$ , as expressed in Eq. (10), is used to capture the existence of any pattern in the estimation errors (if  $S_{D-W}$  is close to two, the errors are essentially random).

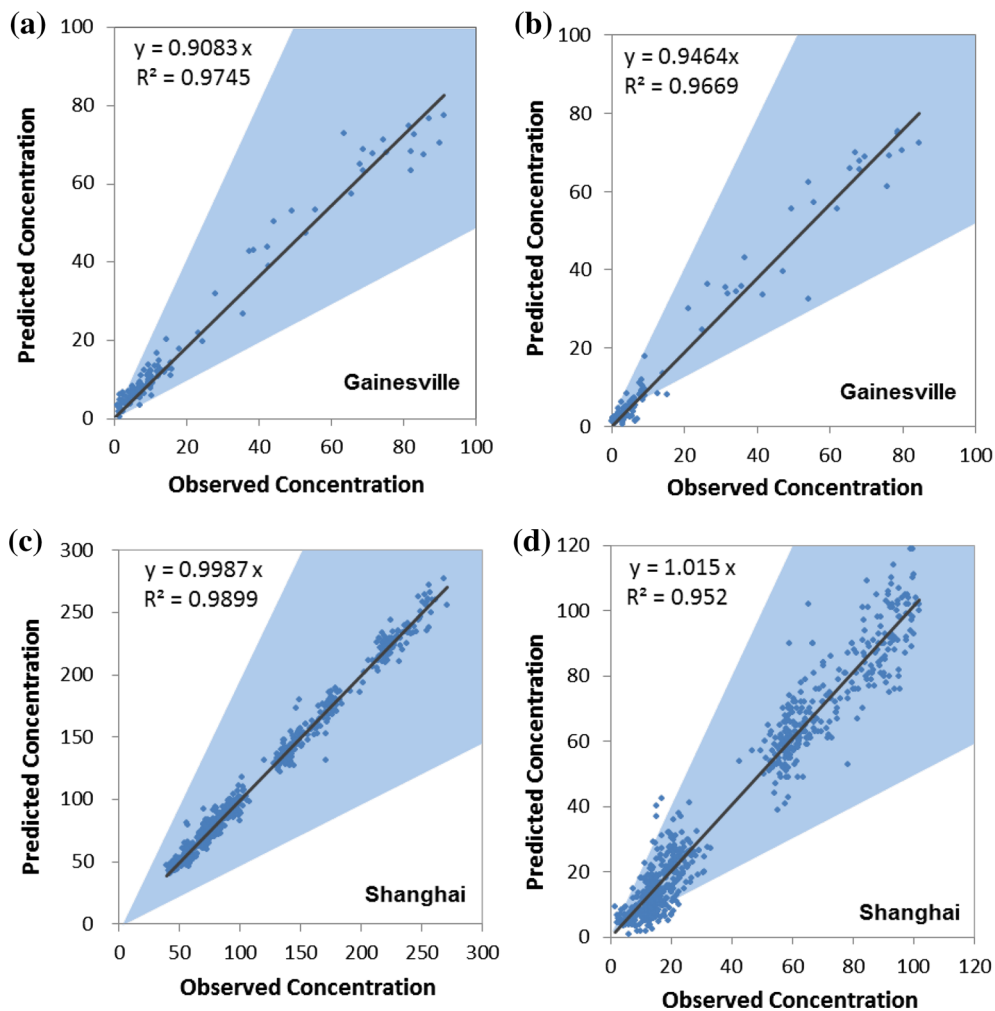
$$T = \frac{\sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}}{\sqrt{\frac{1}{n} \sum_{t=1}^n y_t^2} + \sqrt{\frac{1}{n} \sum_{t=1}^n \hat{y}_t^2}} \quad (9)$$

where,  $0 \leq T \leq 1$ ,  $T \rightarrow 0$  indicates that the model is good for estimation; otherwise, the model is not suitable for estimation.



**Table 3** Performance of different number of hidden neurons applied for ANN-based model

Number of hidden neurons	6	8	9	10	11	12	13	14	16	32
<i>Gainesville US</i>										
MRE (%)	22	36.5	68.1	33.5	25.4	<b>23.6</b>	29	64.8	38.2	36.5
MAE ( $\mu\text{g}/\text{m}^3$ )	2.059	2.301	4.554	2.689	2.067	<b>1.837</b>	2.167	3.237	4.139	3.262
RMSE ( $\mu\text{g}/\text{m}^3$ )	4.008	3.422	6.059	4.027	3.408	<b>2.786</b>	3.989	4.677	2.347	5.042
Number of hidden neurons	4	6	8	9	<b>10</b>	11	12	13	16	32
<i>Shanghai China</i>										
MRE (%)	4.973	5.592	5.069	5.646	<b>4.154</b>	4.352	5.06	4.246	4.936	6.36
MAE ( $\mu\text{g}/\text{m}^3$ )	5.175	5.949	5.164	5.585	<b>4.293</b>	4.408	5.259	4.392	5.248	6.626
RMSE ( $\mu\text{g}/\text{m}^3$ )	7.048	8.065	7.007	7.431	<b>6.084</b>	6.229	7.083	6.097	7.428	8.86

**Fig. 5** Factor-of-two plots (all concentrations in  $\mu\text{g}/\text{m}^3$ ): **a** Gainesville (with background); **b** Gainesville (without background); **c** Shanghai (with background); **d** Shanghai (without background)

$$S_{D-w} = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=2}^n e_t^2} \quad (10)$$

where,  $e_t = y_i - \hat{y}_i$ .

The results indicated that there were improved  $R^2$  and  $S_{D-w}$  for both Gainesville and Shanghai sites after removing the background concentration. The Theil's





**Table 4** Descriptive statistics of ANN-based models for estimation

Statistics measure	Gainesville		Shanghai	
	Scenario 1 <sup>a</sup>	Scenario 2 <sup>b</sup>	Scenario 1	Scenario 2
ME	0.697	0.113	−0.082	0.165
MAE	1.837	2.123	4.293	5.220
RMSE	2.786	3.658	6.084	7.008
$R^2$	0.876	0.936	0.924	0.954
$T$	0.077	0.080	0.023	0.071
$S_{D-W}$	0.709	1.368	1.577	1.604

<sup>a</sup> Scenario 1 indicated the ANN-based model considering the factor of background concentration

<sup>b</sup> Scenario 2 indicated the ANN-based model after removing the background concentration influence

inequality coefficient slightly increased, but is still rather small which indicates a good estimation.  $R^2$  were 0.936 in Gainesville and 0.954 in Shanghai after removing the influence of background concentration. The RMSE showed the same result that model performed better without considering background concentration. The performances in various scenarios are consistent with exploratory analyses above, indicating that background concentration is one of dominant factors influencing the prediction model. Comparing Scenario 2 in Gainesville and Shanghai,  $R^2$  was higher in Shanghai, indicating a better estimation performance when predicting  $PM_{2.5}$  with relative high background concentration.

## Conclusion

In this study, an ANN-based approach was proposed to estimate  $PM_{2.5}$  dispersion near arterials. Different from traditional prediction model, ANN showed capabilities in predicting highly nonlinear relationship. Factors that influenced the dispersion were classified into four aspects, which were traffic-related, weather-related, location-related and background-related, among which background concentrations significantly counts for the ANN estimation.

The monitoring campaign detailed in this study showed that there was an evident concentration decay along the perpendicular direction from the road within 50 m in Gainesville, FL., while in Shanghai the decay gradient was smaller due to influence from high background concentration. In Gainesville's case, 82.2 % of total samples had the concentrations less than  $20 \mu\text{g}/\text{m}^3$ , among which background concentration accounted for 28.9 % of the average detected concentration. The data of monitoring in Shanghai demonstrated 93.6 % of the total samples had the concentration higher than  $50 \mu\text{g}/\text{m}^3$ , and background

concentration accounted for 70.8 % of average concentration detected. Compared with previous researches (Shiva Nagendra and Khare 2004; Cai et al. 2009), this study provided the on-site observed concentrations of  $PM_{2.5}$  within 50 m near arterials. Both local weather condition and traffic volume were recorded minute by minute, which were used for ANN-based prediction and further analyses.

Different structures of models were established based on Gainesville and Shanghai monitoring data. Hidden neurons were determined for both sites by comparing the prediction performance. Besides, both scenarios, with and without background, were considered and analyzed. Results indicated that, ANN-based model predicted more accurately when kicking off the background influence. Meanwhile, the higher increments of the detection point, the better prediction of the model.

This study presents an improvement over previous studies. The portable PM detectors and weather station provide local-scale information, which are more accurate and convincing for usage in local-scale dispersion prediction. However, there are still challenges when applying ANN-based model at a larger scale. (1) The monitor campaign sites in this study were chosen by considering open-type streets without considering the influence of street types. The street canyon effect needs to be checked when applying the ANN-based model to urban area. (2) To improve the model performance, parameters in the input layer still need to be further determined, such as street slope, coverage of vegetation. (3) A large scale of on-site monitoring campaign is necessary and essential for prediction estimation. Longitudinal monitoring campaign which could last for a year or longer is recommended to analyze the inner relationship between  $PM_{2.5}$  dispersion and seasonal variations.

**Acknowledgments** This research was supported and sponsored by the Key Lab of Ocean Engineering of Shanghai Jiao Tong University (Grant Number: JKZD010059). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsors. The authors are also grateful for the strong support and valuable suggestions from Dr. Daniel (Jian) SUN and Dr. Qing-Chang LU, as well as great field data collection assistance from Jingjing CHANG, Yue YU and Tian-Qi ZHANG in the Center for ITS and UAV Applications Research, Shanghai Jiao Tong University.

## References

- Abdullah LC, Wong LL, Saari M, Salmiaton A, Rashid MA (2007) Particulate matter dispersion and haze occurrence potential studies at a local palm oil mill. *Int J Environ Sci Technol* 4(2):271–278
- Annual average daily traffic report (2012) Florida Department of Transportation. <http://www2.dot.state.fl.us/FloridaTrafficOnline>. Accessed June 15, 2013



- Bandyopadhyay G, Chattopadhyay S (2007) Single hidden layer artificial neural network models versus multiple linear regression model in forecasting the time series of total ozone. *Int J Environ Sci Technol* 4:141–149
- Beckerman B, Jerrett M, Brook JR, Verma DK, Arain MA, Finkelstein MM (2008) Correlation of nitrogen dioxide with other traffic pollutants near a major expressway. *Atmos Environ* 42(2):275–290
- Brugge D, Durant JL, Rioux C (2007) Near-highway pollutants in motor vehicle exhaust: a review of epidemiologic evidence of cardiac and pulmonary health risks. *Environ Health* 6(1):23
- Cai M, Yin Y, Xie M (2009) Prediction of hourly air pollutant concentrations near urban arterials using artificial neural network approach. *Transp Res Part D Transp Environ* 14(1):32–41
- Chen H, Bai S, Eisinger D, Niemeier D, Claggett M (2009) Predicting near-road PM 2.5 concentrations. *Transp Res Rec J Transp Res Board* 2123(1):26–37
- Clench-Aas J, Bartnova A, Böhler T, Grønskei KE, Sivertson B, Larssen S (1999) Air pollution exposure monitoring and estimating. Part I: integrated air quality monitoring system. *J Environ Monit* 1(4):313–319
- Gardner MW, Dorling SR (1998) Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmos Environ* 32(14–15):2627–2636
- Hagan MT, Demuth HB, Beale MH (1996) Neural network design. Pws, Boston
- Hitchins J, Morawska L, Wolff R, Gilbert D (2000) Concentrations of submicrometre particles from vehicle emissions near a major road. *Atmos Environ* 34(1):51–59
- Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural netw* 2(5):359–366
- Kam W, Liacos JW, Schauer JJ, Delfino RJ, Sioutas C (2012) On-road emission factors of PM pollutants for light-duty vehicles (LDVs) based on urban street driving conditions. *Atmos Environ* 61:378–386
- Karlaftis MG, Vlahogianni EI (2011) Statistical methods versus neural networks in transportation research: differences, similarities and some insights. *Transp Res Part C Emerg Technol* 19(3):387–399
- Kukkonen J, Partanen L, Karppinen A, Ruuskanen J, Junninen H, Kolehmainen M, Cawley G (2003) Extensive evaluation of neural network models for the prediction of NO<sub>2</sub> and PM<sub>10</sub> concentrations, compared with a deterministic modelling system and measurements in central Helsinki. *Atmos Environ* 37(32):4539–4550
- Lighty JS, Veranth JM, Sarofim AF (2000) Combustion aerosols: factors governing their size and composition and implications to human health. *J Air Waste Manag Assoc* 50(9):1565–1618
- Lowenthal DH, Gertler AW, Labib MW (2014) Particulate matter source apportionment in Cairo: recent measurements and comparison with previous studies. *Int J Environ Sci Technol* 11(3):657–670
- Ministry of Housing and Urban-Rural Construction of the P. R. China (2012) Code for design of urban road engineering, China Architecture and Building Press
- Nagendra SM, Khare M (2006) Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions. *Ecol Model* 190(1):99–115
- National Research Council (2000) Modeling mobile source emissions. National Academies Press, Washington, DC
- Ordieres JB, Vergara EP, Capuz RS, Salazar RE (2005) Neural network prediction model for fine particulate matter (PM<sub>2.5</sub>) on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua). *Environ Model Softw* 20(5):547–559
- Rumelhart DE, McClelland JL, PDP Research Group (1986) Parallel distributed processing: explorations in the microstructure of cognition, Volumes 1 and 2
- Shiva Nagendra SM, Khare M (2004) Artificial neural network based line source models for vehicular exhaust emission predictions of an urban roadway. *Transp Res Part D Transp Environ* 9(3):199–208
- Tiitta P, Raunemaa T, Tissari J, Yli-Tuomi T, Leskinen A, Kukkonen J, Harkonen J, Karppinen A (2002) Measurements and modelling of PM<sub>2.5</sub> concentrations near a major road in Kuopio, Finland. *Atmos Environ* 36(25):4057–4068
- US Department of Energy (2010) Transportation Energy Data Book, Edition 29 of ORNL-5198, Office of Energy Efficiency and Renewable Energy
- Vallamsundar S, Lin JJ (2012) MOVES and AERMOD used for PM 2.5 conformity hot spot air quality modeling. *Transp Res Rec J Transp Res Board* 2270(1):39
- World Health Organization (2002) The world health report 2002: reducing risks, promoting healthy life. World Health Organization, Geneva
- Yura EA, Kear T, Niemeier D (2007) Using CALINE dispersion to assess vehicular PM<sub>2.5</sub> emissions. *Atmos Environ* 41(38):8747–8757
- Zhao WC, Cheng JP, Yu ZY, Tang QL, Cheng F, Yin YW, Wang WH (2013) Levels, seasonal variations, and health risks assessment of ambient air pollutants in the residential areas. *Int J Environ Sci Technol* 10(3):487–494
- Zhu Y, Hinds WC, Kim S, Shen S, Sioutas C (2002) Study of ultrafine particles near a major highway with heavy-duty diesel traffic. *Atmos Environ* 36(27):4323–4335
- Zhu Y, Kuhn T, Mayo P, Hinds WC (2006) Comparison of daytime and nighttime concentration profiles and size distributions of ultrafine particles near a major highway. *Environ Sci Technol* 40(8):2531–2536

