Air Pollution Forecasting Using Artificial and Wavelet Neural Networks with Meteorological Conditions

Qingchun Guo^{1,2*}, Zhenfang He^{1,3*}, Shanshan Li¹, Xinzhou Li^{2,4}, Jingjing Meng¹, Zhanfang Hou¹, Jiazhen Liu¹, Yongjin Chen¹

¹ School of Environment and Planning, Liaocheng University, Liaocheng 252000, China

² State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

³ State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

⁴ CAS Center for Excellence in Tibetan Plateau Earth Sciences, Beijing 100101, China

ABSTRACT

Air quality forecasting is a significant method of protecting public health because it provides early warning of harmful air pollutants. In this study, we used correlation analysis and artificial neural networks (ANNs; including wavelet ANNs [WANNs]) to identify the linear and nonlinear associations, respectively, between the air pollution index (API) and meteorological variables in Xi'an and Lanzhou. Evaluating twelve algorithms and nineteen network topologies for the ANN and WANN models, we discovered that the optimal input variables for an API forecasting model were the APIs from the 3 preceding days and sixteen selected meteorological factors. Additionally, the API could be accurately predicted based solely on the value recorded 3 days earlier. Based on the correlation coefficients between the air pollution index of the targeted day and the tested variables, the API displayed the closest relationship with the API 1 day earlier as well as stronger correlations with the average temperature, average water vapor pressure, minimum temperature, maximum temperature, API 2 days earlier, and API 3 days earlier. When Bayesian regularization was applied as a training algorithm, the WANN and ANN models accurately reproduced the APIs in both Xi'an and Lanzhou, although the WANN model ($R = 0.8846$ for Xi'an and $R = 0.8906$ for Lanzhou) performed better than the ANN ($R = 0.8037$ for Xi'an and $R = 0.7742$ for Lanzhou) during the forecasting stage. These results demonstrate that WANNs are effective in short-term API forecasting because they can recognize historic patterns and thereby identify nonlinear relationships between the input and output variables. Thus, our study may provide a theoretical basis for environmental management policies.

Keywords: Air pollution; Wavelet artificial neural network; Meteorological factor; Forecast.

INTRODUCTION

Air pollution is a theme of high importance, and global problems have demonstrated its damaging impacts on human physical health and ecosystems (Nguyen *et al.*, 2015). Meanwhile, it also has a detrimental effect on visibility, climate, and sustainable development (Lelieveld *et al.*, 2015). Poor air quality is one of the five major health risks in the world, for example, long-term exposure to polluted air is related to respiratory infections, heart attack, stroke and lung cancer (Kessler, 2014; Watson, 2014; Lelieveld and Pöschl, 2017). And air pollution has adverse effects on

people's life span, and social communication willingness (Huang *et al.*, 2018).

Due to the large-scale development of industrialization and urbanization, China has been suffering from acute air pollution for many years (Liu and Diamond, 2005). The number of haze days in a year has also risen evidently in China, which has seriously hindered the sustainable development of society and caused widespread concern from all walks of life (Jiang and Bai, 2018). In 2013, China suffered extremely serious haze pollution, influencing 800 million people, and daily average PM_{2.5} concentrations at a site in Xi'an were more than twice those of Beijing, Shanghai, and Guangzhou (Huang *et al.*, 2014).

Air pollution forecasting also is crucial for public health interventions and air pollution control policymaking. However, air quality forecasting is quite complex (Li *et al.*, 2017a; Park *et al.*, 2018). Apart from the rapid economic growth, air pollution is affected by unfavorable meteorological

* Corresponding author.

E-mail address: guoqingchun@lcu.edu.cn (Q. Guo);
hezhenfang@lcu.edu.cn (Z. He)

conditions (Al-Saadi *et al.*, 2005; Gong and Ordieres-Meré, 2016; Li *et al.*, 2017b).

Artificial neural network (ANN) has been performed to predict ground motion (Wiszniowski, 2016), and groundwater depth (He *et al.*, 2014). In particular, ANN has been shown to be effective for more complex tasks. And ANN models utilize a sophisticated technique that has been successfully applied to forecast air pollution (Li *et al.*, 2017b). However, in some cases, the data is too complex for the modeling tools to be processed. Hence it is necessary to preprocess the input system information (Simons *et al.*, 1995). Traditionally, this has been done by using principal component analysis (Xia *et al.*, 2015), or by using Fourier transform (Artursson *et al.*, 2002). Whatever technique is used, it must address two objectives: to save a number of relevant information and to reduce the complexity of the input signal (Zhao *et al.*, 2018). Here, the wavelet transformation was employed to extract the important information from the past air pollution index (API) and meteorological factors. The use of wavelet artificial neural network (WANN) as the predictive model is explained by emphasizing the following aspects: (a) the effects of diverse

network parameters and (b) investigation of the capability of WANN model for forecasting next-day air quality (Bai *et al.*, 2016), which offers important guidance to public.

MATERIAL AND METHODS

Study Area and Data Introduction

The two study stations are Xi'an and Lanzhou, both located in China (Fig. 1). API data were gathered at the Environmental Protection Agency, and meteorological data at the Meteorological Bureau.

Fig. 2 shows API from January 2010 to December 2012. API has a periodic law at both sites, and API is larger in the winter and spring, while it is smaller in autumn and summer. And PM_{10} is the primary air pollutant in both cities, therefore, API may represent PM_{10} .

The data series were divided into a training group (January 2010–December 2011), a calibration group (January–June 2012) and a testing group (July–December 2012). The generalization ability of WANN and ANN is tested by cross-correction.

Fig. 1. Location of monitoring stations (Xi'an and Lanzhou).

Fig. 2. Air pollution index (API) from January 2010 to December 2012.

Artificial Neural Network

Fig. 3 provides an architecture of the artificial neural network employed in the study with one node (API) in the output layer and nineteen nodes in the input layer. The input layer consists of nineteen nodes; namely, precipitation (P), extreme wind speed (EWS), extreme wind speed direction (EWSd), average atmospheric pressure (AAP), average wind speed (AWS), average temperature (AT), average water vapor pressure (AWVP), average relative humidity (ARH), sunshine duration (SD), minimum atmospheric pressure (MAP), minimum temperature (MINT), maximum atmospheric pressure (MAP), maximum temperature (MAXT), maximum wind speed (MWS), maximum wind speed direction (MWSd), minimum relative humidity (MRH), air pollution index [API(t)], API(t – 1), and API(t – 2).

Backpropagation is a general approach to train ANNs to minimize the goal function (global error function) (Nunnari et al., 2004). The global error function (F) is computed by utilizing Formula (1):

$$F = \frac{N}{2} \sum (B_i - D_i)^2 \quad (1)$$

where F is the global error function, B_i is the expected output, and D_i is the output of network prediction. The gradient descent technique is employed to adjust the weights of F minimization by using Formula (2) below:

$$\Delta C_{ji} = -\eta \frac{\partial F}{\partial C_{ji}} \quad (2)$$

Fig. 3. Schematic of the artificial neural network used to forecast air pollution index in this study.

where ΔC_{ji} = weight; and η = learning rate.

Wavelet Transformation

The Mallat pyramidal algorithm is used to calculate the discrete wavelet transform coefficient (DWT) (Mallat, 1989a). So, the DWT was employed to analyze the API and meteorological data. The DWT also comprises a multi-resolution decomposition scheme for input signals (Mallat, 1989b). The DWT of a data sequence $f(q)$ is defined as Formula (3):

$$f(u, v) = \frac{1}{\sqrt{u}} \int_{-\infty}^{\infty} f(q) \psi\left(\frac{q-v}{u}\right) dq \quad (3)$$

where $\psi(q)$ indicates the base wavelet of active length q ; u indicates the scale or dilation factor; v indicates the translation in time. For a discrete signal $f(q)$, $f(q) \in S^2(\mathbb{R})$, the DWT is defined by multi-resolution decomposition, which can be calculated by the Mallat decomposition algorithm and Mallat pyramidal reconstruction algorithm (Li et al., 1997):

$$Y_m^0 = f[m], m \in M \quad (4)$$

$$Y_m^i = \sqrt{2} \sum_{l \in \mathbb{Z}} r(l-2m) Y_l^{i-1}, i = 1, 2, 3, \dots, S \quad (5)$$

$$K_m^i = \sqrt{2} \sum_{l \in \mathbb{Z}} t(l-2m) Y_l^{i-1}, i = 1, 2, 3, \dots, S \quad (6)$$

where t and r are the impulse responses to high-pass filter T and low-pass filter R , respectively; K_m^i and Y_m^i are the wavelet series and dimension of the 2^{-i} dimension, respectively; and S is the maximum probable dimension of the discrete data $f[m]$. The Mallat pyramidal reconstruction formula is:

$$Y_m^{i-1} = \sum_{l \in \mathbb{Z}} [\bar{r}(l-2m) Y_l^i + \bar{t}(l-2m) K_m^i] \quad (7)$$

where $\bar{r}$ and $\bar{t}$ are the impulse responses to R^* and T^* , respectively, that is, $R^* = \bar{R}^T$, $T^* = \bar{T}^T$.

The major aim of using DWT is to decrease the complicity of input signal and the number of related information between decomposition compositions (detailed CD1, CD2 and approximate CA2). DWT can be employed to approximate components to obtain low-dimension compositions and get that of multi-dimension analysis. The correlation coefficients of CD1, CA2 and CD2 are less than 0.0037. It turned out to be the best way to achieve our goal.

Wavelet Artificial Neural Network

We use WANN model architecture (Fig. 4) to decompose the original time series (API) into three sets of data: detailed CD2 and CD1 components and approximate CA2. Afterwards, these data are used by the ANN as the input elements. In Fig. 4, A_n is input variables, API_{n+t} is the next-day API.

Fig. 4. Schematic of a wavelet artificial neural network (WANN) used to forecast air pollution index.

Evaluation Criteria

Four performance criteria are employed to assess the validity of WANNs and ANNs adopted in the research. These are root mean square error (RMSE), mean error (ME), percentage error of peak (EO_p), and correlation coefficient (R), which are as follows (He *et al.*, 2014):

$$RMSE = \sqrt{\frac{\sum (B_j - D_j)^2}{L}} \quad (8)$$

$$ME = \frac{1}{L} \sum (B_j - D_j) \quad (9)$$

$$EO_p = \frac{D_p - B_p}{B_p} \times 100\% \quad (10)$$

$$R = \frac{\frac{1}{L} \sum (B_j - \bar{B})(D_j - \bar{D})}{\sqrt{\frac{1}{L} \sum (B_j - \bar{B})^2} \sqrt{\frac{1}{L} \sum (D_j - \bar{D})^2}} \quad (11)$$

where B_j = measured API for the j^{th} data, D_j = fitting API for the j^{th} data, $\bar{B}$ = mean of measured API, $\bar{D}$ = mean of fitting API, L = number of measures, D_p expresses the peak of the fitting API, B_p is the peak of the measured API and EO_p is the relative error of peak API.

RESULTS AND DISCUSSION

Correlation Analysis

Correlation analysis can determine the linear associations between air pollution and meteorological variables. The main disadvantage of using correlation analysis is that it could only detect the linear relationship between two variables. As a result, correlation analysis cannot catch any

possible nonlinear relationship that may exist between the outputs and the inputs and may result in missing important output-related inputs in a nonlinear fashion.

The determination of input parameters is one of the most important steps in the design of WANN models. The selection of correlation functions calculated for the variables is shown in Table 1, which is 95% significant. The performance of every variable was evaluated by computing its correlation coefficient (R) with $API(t+1)$. The analysis showed that $API(t)$ was strongly related to $API(t+1)$ at the two stations. Furthermore, the performances of average temperature, average water vapor pressure, minimum temperature, maximum temperature, $API(t-1)$, and $API(t-2)$ were better than other variables at two stations. That is, the meteorological parameters with the highest correlation to $API(t+1)$ comprise the above variables. We identified seven significant variables. Therefore, different combinations of variables were selected as inputs for modeling daily API in Table 2. The selection of the variables was based both on comprehensive correlation analysis and on existing knowledge. The horizontal wind is the basic parameter that controls the horizontal dispersion and transport of air pollutants. The effects of solar radiation on the reaction rate constants and, consequently, on the destruction and formation of photochemical species, are complicated. The removal of air pollutants from the atmosphere by precipitation is a very effective process that often leads to low air pollution levels. Many pollutants are highly persistent, and it is usually accepted that the possibility of occurrence of air pollution events increase if the past day's air pollution was higher than normal.

Determination of Network Topologies and Training Algorithms

It should be emphasized that finding the most appropriate model structure may be one of the main tasks of the model developer. That is probably because there are usually a lot of candidate variables, and the priority is unknown. Moreover, the relationship between inputs and air quality is nonlinear and highly location dependent.

Table 1. Correlation coefficient (R) between API(t + 1) and input variables.

Number	Influence factor	Abbreviation	R	
			Xi'an	Lanzhou
1	Precipitation (t)	P(t)	−0.2772	−0.1775
2	Extreme wind speed (t)	EWS(t)	−0.1996	−0.029
3	Extreme wind speed direction (t)	EWSD(t)	−0.0542	−0.0455
4	Average atmospheric pressure (t)	AAP(t)	0.2003	0.067
5	Average wind speed (t)	AWS(t)	−0.1737	0.0274
6	Average temperature (t)	AT(t)	−0.3526	−0.3114
7	Average water vapor pressure (t)	AWVP(t)	−0.3853	−0.3763
8	Average relative humidity (t)	ARH(t)	−0.1657	−0.1959
9	Sunshine duration (t)	SD(t)	−0.0414	0.1217
10	Minimum atmospheric pressure (t)	MAP(t)	0.1975	0.033
11	Minimum temperature (t)	MINT(t)	−0.3837	−0.3589
12	Maximum atmospheric pressure (t)	MAP(t)	0.2137	0.108
13	Maximum temperature (t)	MAXT(t)	−0.3225	−0.227
14	Maximum wind speed (t)	MWS(t)	−0.1648	−0.006
15	Maximum wind speed direction (t)	WSD(t)	−0.0225	0.0096
16	Minimum relative humidity (t)	MRH(t)	−0.1521	−0.2465
17	Air pollution index (t)	API(t)	0.6597	0.5126
18	Air pollution index (t − 1)	API(t − 1)	0.4709	0.3248
19	Air pollution index (t − 2)	API(t − 2)	0.3357	0.2444

Table 2. Input variables for the ANN and WANN models.

Model ID	Input variables	Architecture
ANNAPI1	P(t), EWS(t), EWSD(t), AAP(t), AWS(t), AT(t), AWVP(t), ARH(t), SD(t), MAP(t), MINT(t), MAP(t), MAXT(t), MWS(t), MWSD(t), MRH(t), API(t), API(t − 1), API(t − 2)	19:3:1
WANNAPI1	P(t), EWS(t), EWSD(t), AAP(t), AWS(t), AT(t), AWVP(t), ARH(t), SD(t), MAP(t), MINT(t), MAP(t), MAXT(t), MWS(t), MWSD(t), MRH(t), API(t), API(t − 1), API(t − 2)	57:6:1
WANNAPI2	P(t), EWS(t), EWSD(t), AAP(t), AWS(t), AT(t), AWVP(t), ARH(t), SD(t), MAP(t), MINT(t), MAP(t), MAXT(t), MWS(t), MWSD(t), MRH(t), API(t)	51:6:1
WANNAPI3	P(t), EWS(t), EWSD(t), AAP(t), AT(t), AWVP(t), ARH(t), SD(t), API(t), API(t − 1), API(t − 2)	33:6:1
WANNAPI4	AT(t), AWVP(t), MINT(t), MAXT(t), API(t), API(t − 1), API(t − 2)	21:6:1
WANNAPI5	AT(t), AWVP(t), MINT(t), MAXT(t)	12:6:1
WANNAPI6	AWVP(t), MINT(t)	6:6:1
WANNAPI7	API(t)	3:6:1
WANNAPI8	API(t), API(t − 1), API(t − 2)	9:6:1

The different details and dimensions of the input variables are obtained by two-stage decomposition of the wavelet transform. After two-stage decomposition and reconstruction, the variables are divided into three portions. The approximate composition CA2 indicates the general trend of the original variables, while the detailed CD2 reflects the periodic values of the original variables, and the detailed CD1 reflects the inhomogeneity and complicity of the original variables. In other words, the detailed CD1 determines the complexity of empirical model predictions.

The variation characteristics of sequences are the critical elements affecting the selection of wavelets (Sang, 2013). In order to decompose the input variables optimally, the mother wavelet is selected, and the similarity between the CD1, CD2 and CA2 is considered. The minimum R can best satisfy our purpose of analyzing the variation characteristics of different components of the input variables. The quantitative calculation shows that the components are independent of

each other. Twenty-one kinds of wavelet functions are selected for DWT. Table 3 shows that db4 is the best wavelet function in the study because it has the smallest R. Here we take the average temperature as an example, and other input variables have similar results.

Trial and error is applied to acquire the optimal model parameters. Fig. 5 shows that network topologies (19-3-1 for ANN and 57-6-1 for WANN) for Xi'an are better than others by trial and error. The amount of nodes of the hidden layer rises from 1 to 19 in the models. The following observations can be made from Fig. 5 as raising the amount of nodes of the hidden layer: The RMSE values decreased slightly, but after 19-3-1 for ANN and 57-6-1 for WANN, RMSE values increase and fluctuate. Therefore, the best topologies of the patterns for Xi'an are separately identified as 19-3-1 for ANN and 57-6-1 for WANN. Similarly, the best system topologies for Lanzhou are separately 19-3-1 for ANN and 57-6-1 for WANN.

Table 3. Correlation coefficient (R) between CA2, CD1 and CD2 for various wavelet functions in Xi'an.

Wavelet function	R		
	CA2 and CD1	CA2 and CD2	CD1 and CD2
db2	0.0015	−0.0013	0
db3	0.0013	−0.0007	0.0007
db4	0.0001	0.0007	0
db5	−0.0009	0.0007	−0.0007
db6	−0.0011	0.0006	−0.0016
db7	−0.0005	0	−0.0007
db8	0.0004	−0.0001	−0.0001
sym2	0.0015	−0.0013	0
sym4	−0.0014	0.0005	0.0006
sym6	−0.001	0.001	−0.0011
sym8	−0.001	0.0005	−0.0004
coif1	−0.0017	0.0008	0.0004
coif2	−0.0015	0.0011	0.0001
coif3	−0.0012	0.0009	−0.0004
coif4	−0.001	0.0007	−0.0007
bior1.1	0.0008	0.0007	0.0005
bior2.2	0.0025	0.0034	0.0208
bior3.3	0.0018	−0.0071	0.0132
bior4.4	−0.0017	0.0048	0.0202
bior5.5	0.0015	0.0002	0.0091
bior6.8	−0.0011	0.0022	0.0039

Fig. 5. Optimization of network topologies and training algorithms in Xi'an.

Fig. 5 shows the performances of the improved training algorithms, revealing that the trainbr algorithm has the best performance in predicting $API(t + 1)$ in Xi'an. Trainbr automatically sets optimum values for the parameters of the objective function.

Table 4 shows transfer function (tansig-purelin) in Xi'an is better than others during training, cross-validation and testing periods. Similarly, transfer function (tansig-purelin) in Lanzhou is also better than others.

Comparative Analysis of the Models

All results of trainings for ANN and WANN during the training period are shown in Table 5. The RMSE for the ANN ANNAPI1 and WANNAPI1 in Xi'an are 22.7233 and 11.5683, respectively; the R are 0.642 and 0.9783, respectively; the ME are −0.1114 and 0.7682, respectively; and the EO_p are −0.5337% and −0.016%. The WANNs are superior to the ANN during the training period. Meanwhile, similar results in Lanzhou also can be found in the RMSE,

ME, R, EO_p .

The values of the evaluative criteria for the nine models at the two stations during the forecast period are shown in Table 6. Table 6 summarizes the results of the tests with every network configuration. The ANNs and WANNs have an agile mathematic structure and can map highly nonlinear relations. Most WANN models have good performance in Xi'an and Lanzhou. However, the performances of WANN models in Xi'an were obviously superior to those of the WANN models in Lanzhou. The performance of WANNAPI1 is better than that of ANNAPI1 in Xi'an and Lanzhou. The WANNAPI1 model had even more obvious advantages for Lanzhou, where it was found to provide a more accurate API forecast than the ANNAPI1 model. The EO_p values in Table 6 show the models' performances in simulating the extreme events. In Lanzhou and Xi'an, during the forecast period, the RMSE value of the WANNAPI1 model was the smallest of all the WANN models; the R-value of WANNAPI1 was the largest, and EO_p was the smallest. The lower RMSE values

Table 4. Comparison between various transfer functions in Xi'an.

Transfer function	ANN			WANN		
	RMSE (training)	RMSE (validation)	RMSE (testing)	RMSE (training)	RMSE (validation)	RMSE (testing)
tansig-purelin	21.0461	28.6731	27.7817	8.2565	25.7647	25.2995
tansig-logsig	165.639	133.9647	166.0429	165.4297	133.2572	165.986
tansig-tansig	21.6122	30.9392	40.8472	6.4052	30.7811	30.657
logsig-purelin	21.5672	29.2643	32.411	10.1233	30.1516	31.745
logsig-tansig	21.0549	29.2643	37.0629	5.7477	29.1385	36.0737
logsig-logsig	165.639	133.8559	166.0429	165.6389	133.1871	166.0429
purelin-tansig	22.6304	29.9151	32.594	10.8048	29.7687	31.0784
purelin-logsig	165.639	140.1284	166.1619	165.6317	138.1583	166.0467
purelin-purelin	22.0284	29.0213	31.0085	11.6997	29.6912	31.7083

Table 5. Comparison of performance statistics during the training period.

Model	Xi'an				Lanzhou			
	RMSE	ME	R	EO _p (%)	RMSE	ME	R	EO _p (%)
ANNAPI1	22.7233	−0.1114	0.642	−0.5337	47.3539	3.5148	0.6782	−0.3887
WANNAPI1	11.5683	0.7682	0.9783	−0.016	19.9275	−2.7785	0.9515	0.0816
WANNAPI2	10.0819	0.8584	0.9412	−0.1333	26.4906	1.2632	0.9112	0.2073
WANNAPI3	7.6149	0.5388	0.9666	−0.0373	21.0269	0.4052	0.9448	0.0936
WANNAPI4	11.4359	0.212	0.9222	−0.3981	19.7556	−0.5168	0.9515	0.0443
WANNAPI5	26.3829	0.084	0.519	−0.734	54.6691	−0.1092	0.5233	−0.6211
WANNAPI6	26.0514	0.808	0.4735	−0.7578	57.8112	−1.5231	0.4362	−0.7169
WANNAPI7	17.5568	0.8278	0.8065	−0.5349	39.7627	0.338	0.7847	−0.3203
WANNAPI8	11.904	0.7964	0.9157	−0.417	20.4221	0.601	0.948	0.0407

Table 6. Comparison of performance statistics during the forecast period.

Model	Xi'an				Lanzhou			
	RMSE	ME	R	EO _p (%)	RMSE	ME	R	EO _p (%)
ANNAPI1	27.7817	5.4455	0.8037	−21.69	53.8214	36.9678	0.7742	51.9
WANNAPI1	25.2995	6.5498	0.8846	−1.6	28.0526	−15.1728	0.8906	0.0195
WANNAPI2	25.9907	−6.6528	0.8491	−13.33	71.9634	21.9724	0.7685	20.73
WANNAPI3	29.0921	6.3298	0.8836	−3.73	60.2835	−2.3163	0.8584	9.36
WANNAPI4	28.4928	5.1264	0.8608	−39.81	64.4321	23.4622	0.8598	4.43
WANNAPI5	43.4797	−8.0227	0.3581	−73.4	29.6079	17.3015	0.6586	−62.11
WANNAPI6	42.6505	−9.4939	0.4255	−75.78	28.4364	6.5429	0.5237	−71.69
WANNAPI7	29.5702	−7.4844	0.7852	−53.49	53.7512	16.2583	0.7748	−32.03
WANNAPI8	29.5525	8.7991	0.842	−41.7	64.4042	22.7747	0.844	4.07

indicate that the WANNAPI1 model produced fewer differences and discrepancies between the forecasted API($t + 1$) and observed API($t + 1$).

Fig. 6 shows observed API($t + 1$) versus predicted API($t + 1$) in Xi'an and Lanzhou. ANN and WANN models were both able to replicate average of API, however, limited in capturing minimal or maximal peaks.

Figs. 7 and 8 indicates that the ANN and WANN models predicted API at an acceptable accuracy level in Xi'an and Lanzhou. However, the performances of WANN models were obviously superior to those of the ANN models. The WANN models yielded a good agreement between the observed API($t + 1$) and predicted API($t + 1$), but it is obvious that the WANNAPI1 model was better than the WANNAPI2. It is also obvious that the WANNAPI8 model with 1–3-day lag API was better than the WANNAPI7 with 1-day lag API;

that is to say, including the three previous days' API as parameters in input data set gives more precise results. However, it is necessary to point out that the WANN methods have limitations inherent to their structures.

The agreement between the observed API($t + 1$) and the predicted API($t + 1$) is also very good at both stations using WANNAPI4 model. The main meteorological conditions of air pollution are average temperature (t), average water vapor pressure (t), minimum temperature (t) and maximum temperature (t) in Xi'an and Lanzhou. The possible reason is that correlation coefficient between them and air pollution is larger.

Comparison with Other Models

Many studies have been developed to identify and understand the relationships between air quality and

Fig. 6. Boxplots show the variation of observed API($t+1$) (*Observed-API*) and predicted API($t+1$) (such as P-ANNAPI1) in Xi'an and Lanzhou.

Fig. 7. Comparison between the observed API and the predicted API in Xi'an.

meteorological conditions. ANN, which has the abilities of self-adaption and nonlinear mapping, has been certified in its advantage and widespread application in forecasting air quality. The estimated $\text{PM}_{2.5}$ result in Beijing is in a better RMSE ($= 24.06 \text{ mg m}^{-3}$) using ANN than that obtained

through multi-variate statistical analysis method (RMSE = 26.69 mg m^{-3}) (Ni *et al.*, 2017). In the linear regression analysis, the range of R^2 at six subway stations were 0.18–0.63. Nevertheless, the neural network model with present time variables has high R^2 of 0.54–0.81 (Park *et al.*, 2018).

Fig. 8. Comparison between the observed API and the predicted API in Lanzhou.

The ANNs' results provided absolute errors in predicting PM_{10} , SO_2 , and CO are 35%, 43%, 28%, respectively (Kurt *et al.*, 2008). But for the ANN model of API forecasting, the correlation coefficients are 0.6993, 0.6056, 0.6300 for SO_2 , PM_{10} , NO_2 (Jiang *et al.*, 2004).

WANN has better performance for forecasting SO_2 , PM_{10} , and NO_2 in Chongqing than the ANN, such as, RMSE is lower at 4.447 mg m^{-3} , 8.233 mg m^{-3} , and 2.785 mg m^{-3} , respectively (Bai *et al.*, 2016). The best forecast of $\text{PM}_{2.5}$ in Dingling is completed for next day utilizing the hybrid model combining ANN, wavelet transformation, and air mass trajectory, and RMSE is 15.65 mg m^{-3} . It is also noticed that wavelet transform plays a role improving the $\text{PM}_{2.5}$ forecasting accuracy (Feng *et al.*, 2015).

The simulation and forecast were proved by utilizing the data of $\text{PM}_{2.5}$ in Wuhan based on support vector machine. The results showed that the way can obtain precise outcomes (He *et al.*, 2018). The prediction results of neural networks are better than that of linear model, and the maximum prediction error 21 hours ahead is 32% (Perez and Menares, 2018). The long short-term memory (LSTM) can effectively forecast air pollution and achieve the best results (Karimian *et al.*, 2019). The prediction of the 2016 ozone season using

generalized additive models are in good agreement with the relevant measurement results ($R^2 = 0.70$) (Pernak *et al.*, 2019). The average consistency index between $\text{PM}_{2.5}$ prediction and observation for the four seasons in the Yangtze Delta is between 74% and 77%, using machine learning and WRF (Jia *et al.*, 2019). The best estimation of $\text{PM}_{2.5}$ ($R^2 = 0.84$) is obtained by using artificial neural network (Bai *et al.*, 2020). Compared with WRF, the correlation coefficients of machine learning model are higher by 50–100%, which can provide better $\text{PM}_{2.5}$ prediction (Ma *et al.*, 2020).

These studies have improved the artificial neural network and achieved better predicting results, but it still need to enhance the predicting accuracy. Therefore, learning and characteristic collection of historical data plays an important role in guaranteeing predicting accuracy.

We have reported the prediction results at both stations. In Xi'an, the WANN prediction is successful, simulating well the peaks, and the location and shape of the main peaks are predicted correctly by the model but slightly overestimating the general background API because the main purpose is to simulate the peaks correctly. In Lanzhou, the agreement was similar; the general characteristics of observed API was also successfully reproduced by the WANN. The agreement

obtained in Lanzhou is good with the model accurately forecasting the location and the magnitude of the main peaks.

CONCLUSIONS

This study presents an optimum system for nonlinear modeling of the daily API using ANNs and WANNs. The input variables (meteorological elements and APIs) for the models were defined via correlation analysis, and discrete wavelet transform was employed to decompose the time series of the meteorological conditions into different dimensions, whereby a unique mixed aspect was decomposed into multiple unique aspects. This data was then incorporated into the models to simulate the next day's API. Our results indicated that both the ANN and the WANN models predicted the daily API with acceptable accuracy, but the performance of the latter, which integrated the nonlinear mapping of the ANN as well as the multi-scale analysis of the DWT, was obviously superior.

For future WANNs, we will focus on four aspects. Firstly, the models will address meteorological elements and forecast the API in other locations. Secondly, additional elements will be considered, for instance, the longitude, latitude, land use, topography (simulated with the digital elevation model [DEM]), and population density. Sensitivity analysis will also be conducted in order to select parameters that are more closely related to the API, thus improving the predictive accuracy. Thirdly, the models will be used to predict other complex time series that possess nonlinear and unstable characteristics, such as those for the air quality index (AQI), fine particulate matter (PM_{2.5}), PM₁₀, nitrogen dioxide (NO₂), sulfur dioxide (SO₂), carbon monoxide (CO), and ozone (O₃). Finally, a deep reinforcement learning algorithm based on multi-agent cooperation will be employed for air pollution forecasting, thus providing further insights into the multi-scale spatiotemporal prediction of pollutant concentrations.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (41572150, 41472162, 41702373), Shandong Social Sciences Planning Research Program (18CKPJ34), Shandong Province Higher Educational Humanities and Social Science Program (J18RA196), and State Key Laboratory of Loess and Quaternary Geology Foundation (SKLLQG1907). The authors thank the editor and anonymous reviewers for their valuable comments and proposals, which have helped in improving the quality of our article.

REFERENCE

- Al-Saadi, J., Szykman, J., Pierce, R.B., Kittaka, C., Neil, D., Chu, D.A., Remer, L., Gumley, L., Prins, E., Weinstock, L., MacDonald, C., Wayland, R., Dimmick, F. and Fishman, J. (2005). Improving national air quality forecasts with satellite aerosol observations. *Bull. Am. Meteorol. Soc.* 86: 1249–1261. <https://doi.org/10.1175/BAMS-86-9-1249>
- Artursson, T., Spangeus, P. and Holmberg, M. (2002). Variable reduction on electronic tongue data. *Anal. Chim. Acta* 452: 255–264. [https://doi.org/10.1016/S0003-2670\(01\)01448-9](https://doi.org/10.1016/S0003-2670(01)01448-9)
- Bai, L., Huang, L., Wang, Z., Ying, Q., Zheng, J., Shi, X. and Hu, J. (2020). Long-term field evaluation of low-cost particulate matter sensors in Nanjing. *Aerosol Air Qual. Res.* 20: 242–253. <https://doi.org/10.4209/aaqr.2018.11.0424>
- Bai, Y., Li, Y., Wang, X., Xie, J. and Li, C. (2016). Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmos. Pollut. Res.* 7: 557–566. <https://doi.org/10.1016/j.apr.2016.01.004>
- Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L. and Wang, J. (2015). Artificial neural networks forecasting of PM_{2.5} pollution using air mass trajectory based geographic model and wavelet transformation. *Atmos. Environ.* 107: 118–128. <https://doi.org/10.1016/j.atmosenv.2015.02.030>
- Gong, B. and Ordieres-Meré, J. (2016). Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong. *Environ. Modell. Software* 84: 290–303. <https://doi.org/10.1016/j.envsoft.2016.06.020>
- He, P., Zheng, B. and Zheng, J. (2018). Urban PM_{2.5} diffusion analysis based on the improved Gaussian smoke plume model and support vector machine. *Aerosol Air Qual. Res.* 18: 3177–3186. <https://doi.org/10.4209/aaqr.2017.06.0223>
- He, Z., Zhang, Y., Guo, Q. and Zhao, X. (2014). Comparative study of artificial neural networks and wavelet artificial neural networks for groundwater depth data forecasting with various curve fractal dimensions. *Water Resour. Manage.* 28: 5297–5317. <https://doi.org/10.1007/s11269-014-0802-0>
- Huang, J., Pan, X., Guo, X. and Li, G. (2018). Impacts of air pollution wave on years of life lost: A crucial way to communicate the health risks of air pollution to the public. *Environ. Int.* 113: 42–49. <https://doi.org/10.1016/j.envint.2018.01.022>
- Huang, R.J., Zhang, Y., Bozzetti, C., Ho, K.F., Cao, J.J., Han, Y., Daellenbach K.R., Slowik, J.G., Platt, S.M., Canonaco, F., Zotter, P., Wolf, R., Pieber, S.M., Beuns, E.A., Crippa, M., Ciarelli, G., Piazzalunga, A., Schwikowski, M., Abbaszade, G., ... Prévôt, A.S.H. (2014). High secondary aerosol contribution to particulate pollution during haze events in china. *Nature* 514: 218–222. <https://doi.org/10.1038/nature13774>
- Jia, M., Cheng, X., Zhao, T., Yin, C., Zhang, X., Wu, X., Wang, L. and Zhang, R. (2019). Regional air quality forecast using a machine learning method and the WRF model over the Yangtze River Delta, east China. *Aerosol Air Qual. Res.* 19: 1602–1613. <https://doi.org/10.4209/aaqr.2019.05.0275>
- Jiang, D., Zhang, Y., Hu, X., Zeng, Y., Tan, J. and Shao, D. (2004). Progress in developing an ANN model for air pollution index forecast. *Atmos. Environ.* 38: 7055–7064. <https://doi.org/10.1016/j.atmosenv.2003.10.066>
- Jiang, L. and Bai, L. (2018). Spatio-temporal characteristics

- of urban air pollutions and their causal relationships: Evidence from Beijing and its neighboring cities. *Sci. Rep.* 8: 1279. <https://doi.org/10.1038/s41598-017-18107-1>
- Karimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., Zhang, X. and Sachdeva, S. (2019). Evaluation of different machine learning approaches to forecasting PM_{2.5} mass concentrations. *Aerosol Air Qual. Res.* 19: 1400–1410. <https://doi.org/10.4209/aaqr.2018.12.0450>
- Kessler, R. (2014). Prevention: Air of danger. *Nature* 509: S62–S63. <https://doi.org/10.1038/509S62a>
- Kurt, A., Gulbagci, B., Karaca, F. and Alagha, O. (2008). An online air pollution forecasting system using neural networks. *Environ. Int.* 34: 592–598. <https://doi.org/10.1016/j.envint.2007.12.020>
- Lelieveld, J., Evans, J.S., Fnais, M., Giannadaki, D. and Pozzer, A. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature* 525: 367–371. <https://doi.org/10.1038/nature15371>
- Lelieveld, J. and Pöschl, U. (2017). Chemists can help to solve the air-pollution health crisis. *Nature* 551: 291. <https://doi.org/10.1038/d41586-017-05906-9>
- Li, T., Shen, H., Zeng, C., Yuan, Q. and Zhang, L. (2017a). Point-surface fusion of station measurements and satellite observations for mapping PM_{2.5} distribution in China: Methods and assessment. *Atmos. Environ.* 152: 477–489. <https://doi.org/10.1016/j.atmosenv.2017.01.004>
- Li, X., Li, H., Wang, F. and Ding, J. (1997). A remark on the mallat pyramidal algorithm of wavelet analysis wavelet analysis. *Commun. Nonlinear Sci. Numer. Simul.* 2: 240–243. [https://doi.org/10.1016/S1007-5704\(97\)90010-1](https://doi.org/10.1016/S1007-5704(97)90010-1)
- Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C. and Chi, T. (2017b). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environ. Pollut.* 231: 997–1004. <https://doi.org/10.1016/j.envpol.2017.08.114>
- Liu, J. and Diamond, J. (2005). China's environment in a globalizing world. *Nature* 435: 1179–1186. <https://doi.org/10.1038/4351179a>
- Ma, J., Yu, Z., Qu, Y., Xu, J. and Cao, Y. (2020). Application of the XGBoost machine learning method in PM_{2.5} prediction: A case study of Shanghai. *Aerosol Air Qual. Res.* 20: 128–138. <https://doi.org/10.4209/aaqr.2019.08.0408>
- Mallat, S.G. (1989a). Multifrequency channel decompositions of images and wavelet models. *IEEE Trans. Acoust. Speech Signal Process.* 37: 2091–2110. <https://doi.org/10.1109/29.45554>
- Mallat, S.G. (1989b). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 11: 674–693. <https://doi.org/10.1109/34.192463>
- Nguyen, T.T.N., Bui, H.Q., Pham, H.V., Luu, H.V., Man, C.D., Pham, H.N., Le, H.T. and Nguyen, T.T. (2015). Particulate matter concentration mapping from MODIS satellite data: A Vietnamese case study. *Environ. Res. Lett.* 10: 095016. <https://doi.org/10.1088/1748-9326/10/9/095016>
- Ni, X.Y., Huang, H. and Du, W.P. (2017). Relevance analysis and short-term prediction of PM_{2.5} concentrations in Beijing based on multi-source data. *Atmos. Environ.* 150: 146–161. <https://doi.org/10.1016/j.atmosenv.2016.11.054>
- Nunnari, G., Dorling, S., Schlink, U., Cawley, G., Foxall, R. and Chatterton, T. (2004). Modelling SO₂ concentration at a point with statistical approaches. *Environ. Modell. Software* 19: 887–905. <https://doi.org/10.1016/j.envsoft.2003.10.003>
- Park, S., Kim, M., Kim, M., Namgung, H.G., Kim, K.T., Cho, K.H. and Kwon, S.B. (2018). Predicting PM₁₀ concentration in Seoul metropolitan subway stations using artificial neural network (ANN). *J. Hazard. Mater.* 341: 75–82. <https://doi.org/10.1016/j.jhazmat.2017.07.050>
- Perez, P. and Menares, C. (2018). Forecasting of hourly PM_{2.5} in south-west zone in Santiago de Chile. *Aerosol Air Qual. Res.* 18: 2666–2679. <https://doi.org/10.4209/aaqr.2018.01.0029>
- Pernak, R., Alvarado, M., Lonsdale, C., Mountain, M., Hegarty, J. and Nehrkorn, T. (2019). Forecasting surface O₃ in Texas urban areas using random forest and generalized additive models. *Aerosol Air Qual. Res.* 19: 2815–2826. <https://doi.org/10.4209/aaqr.2018.12.0464>
- Sang, Y.F. (2013). A review on the applications of wavelet transform in hydrology time series analysis. *Atmos. Res.* 122: 8–15. <https://doi.org/10.1016/j.atmosres.2012.11.003>
- Simons, J., Bos, M. and Van der Linden, W.E. (1995). Data processing for amperometric signals. *Analyst* 120: 1009–1012. <https://doi.org/10.1039/AN9952001009>
- Watson, T. (2014). Environment: Breathing trouble. *Nature* 513: S14–S15. <https://doi.org/10.1038/513S14a>
- Wiszniowski, J. (2016). Applying the general regression neural network to ground motion prediction equations of induced events in the Legnica-Głogów Copper District in Poland. *Acta Geophys.* 64: 2430–2448. <https://doi.org/10.1515/acgeo-2016-0104>
- Xia, J., Park, J.H. and Zeng, H. (2015). Improved delay-dependent robust stability analysis for neutral-type uncertain neural networks with markovian jumping parameters and time-varying delays. *Neurocomputing* 149: 1198–1205. <https://doi.org/10.1016/j.neucom.2014.09.008>
- Zhao, Y., Zhou, D. and Yan, H. (2018). An improved retrieval method of atmospheric parameter profiles based on the BP neural network. *Atmos. Res.* 213: 389–397. <https://doi.org/10.1016/j.atmosres.2018.06.025>

Received for review, May 6, 2020

Revised, May 21, 2020

Accepted, May 22, 2020