ISSN 2616-5872 Vol.3, Issue 4: 5-10, DOI: 10.25236/AJEE.2021.030402

Research on Air Pollutant Concentration Change Based on Gray Correlation Snalysis and WNN

Yuting He, Shuqi Guo, Tiantian Shi

School of Physics and Electronic-Electrical Engineering, Ningxia University, Yinchuan, Ningxia 750021, China

Abstract: *Air pollution endangers human comfort, health and welfare or the environment. The practice of pollution prevention and control shows that establishing air quality prediction model, knowing the possible air pollution process in advance and taking corresponding control measures are one of the effective methods to reduce the harm of air pollution to human health and environment and improve ambient air quality. At present, WRF-CMAQ simulation system is commonly used to predict air quality, and the effect is not ideal. Therefore, firstly, this paper uses the grey correlation analysis model to study the correlation modeling between the variation characteristics of pollutant concentration and meteorological factors, and reasonably analyzes the meteorological conditions. Through the establishment of wavelet neural network prediction model, the pollutant concentration data are quantitatively analyzed, and then the single day concentration value of conventional pollutants in the future is predicted. The results show that WNN model is feasible to predict the concentration of air pollutants.*

Keywords: *AQI, gray correlation, wavelet neural network, air pollution prediction*

1. Introduction

Atmosphere refers to the air around the earth and is one of the components of the natural environment on which human beings depend [1]. With the rapid increase of the total population of the earth and the continuous development of industry, it has brought a large number of polluting gases and even harmful gas emissions, which makes the human living environment not optimistic. Therefore, it is very necessary to strengthen the monitoring and prediction of atmospheric quality. At present, the monitoring contents of atmospheric quality all over the world mainly include the concentrations of SO_2, NO_2, O_3, CO and suspended particulate matter in the atmosphere. At present, wrf-cmaq simulation system is commonly used to predict air quality. However, due to the uncertainty of simulated meteorological field and emission inventory, as well as the incompleteness of the generation mechanism of pollutants including ozone, the results of wrf-cmaq prediction model are not ideal. Ozone pollution occurs frequently in many regions of the country [2]. The early warning and prevention of ozone pollution is the focus of the environmental protection department. Ozone concentration prediction is also a difficult one among the six pollutants. The reason is that, as the only secondary pollutant among the six pollutants, ozone is not directly discharged from the pollution source, but generated through a series of chemical and photochemical reactions in the atmosphere, which makes it very difficult to accurately predict the change of ozone concentration with WRF-CMAQ model. Therefore, the concept of secondary modeling is proposed, that is, based on the simulation results of primary prediction models such as wrf-cmaq, re modeling is combined with more data sources to improve the accuracy of prediction. How to use the existing measured data and primary prediction data to establish a secondary model to improve the accuracy of ozone prediction is one of the key and difficult points [3].

2. Air quality evaluation based on AQI calculation model

2.1. AQI calculation model

According to the air quality evaluation standard implemented in China, the air quality can be divided into six levels, and the classification standard is determined by the air quality index (AQI). The value range of AQI is 0 to 500. The larger the index, the higher the corresponding air quality level and the more serious the air pollution; On the contrary, the lower the degree of air pollution, the better the air quality. According to the national standards currently implemented in China, the calculated value of AQI is mainly determined by the measured concentration values of various pollutants such as fine particulate matter ($PM_{2.5}$), inhalable particulate matter (PM_{10}), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), ozone (O_3), and carbon monoxide (CO) [4].

According to the classification concentration limit of each item (GB3095-2012), the air quality sub index (IAQIP) corresponding to each pollutant can be calculated by the following formula:

$$\mathrm{IAQI_P} = \frac{\mathrm{IAQI_{Hi}} - \mathrm{IAQI_{Lo}}}{\mathrm{BP_{Hi}} - \mathrm{BP_{Lo}}} \cdot \left(\mathrm{C_P} - \mathrm{BP_{Lo}}\right) + \mathrm{IAQI_{Lo}} \tag{1}$$

Where $IAQI_P$ refers to the air quality sub index of pollutant item P, C_P is the mass concentration value of pollutant item P, and BP_{Hi} is the high value of pollutant concentration limit close to C_P in the air quality sub index of corresponding regions and the corresponding pollutant item concentration index table, BP_{Lo} is the low value of the pollutant concentration limit close to C_P in the air quality sub index of the corresponding region and the corresponding pollutant project concentration index table, and $IAQI_{Hi}$ is the air quality sub index corresponding to BP_{Hi} in the air quality sub index of the corresponding region and the corresponding pollutant project concentration index table, $IAQI_{Lo}$ is the air quality sub index corresponding to BP_{Lo} in the table of air quality sub index of corresponding regions and corresponding pollutant project concentration index.

The maximum 8-hour moving average of ozone (O_3) is calculated as follows:

$$C_{O_3} = \max_{t=8,9,\ldots,24} \left\{ \frac{1}{8} \sum_{i=t-7}^{t} c_t \right\} \tag{2}$$

Where C_t is the average pollutant concentration of ozone from t-1 to t on a certain day.

After calculating the air quality sub index (IAQIP) corresponding to each pollutant, the final air quality index (AQI) is taken as the maximum value of each sub index, namely:

$$\mathrm{AQI} = \max\left\{\mathrm{IAQI_{SO2}}, \mathrm{IAQI_{NO2}}, \mathrm{IAQI_{PM10}}, \mathrm{IAQI_{PM2.5}}, \mathrm{IAQI_{O3}}, \mathrm{IAQI_{CO}}\right\} \tag{3}$$

2.2. Air quality classification based on AQI

The air quality grade range is divided according to the AQI value, and the AQI range corresponding to the grade is shown in Table 1.

Table 1: Air quality grade and corresponding air quality index (AQI) range.

Air quality grade	excellent	good	light pollution	moderate pollution	severe pollution	serious pollution
AQI range	[0,50]	[51,100]	[101,150]	[151,200]	[201,300]	[301,+∞)

3. Correlation analysis between variation characteristics of air pollutant concentration and meteorological factors

Air pollution is mainly affected by two factors, one is the distribution and emission of various pollution sources, and the other is the impact of meteorological factors on air pollutants. The impact of meteorological factors on air quality is mainly reflected in the changes of physical and chemical properties of air pollution. Differences in meteorological conditions lead to differences in pollutant diffusion, dilution and migration [5].

3.1. Grey correlation analysis

Grey correlation analysis refers to the method of quantitative description and comparison of the development and change trend of a system. Its basic idea is to judge whether the relationship is close by determining the geometric similarity between the reference data column and several comparison data columns. It reflects the correlation degree between curves. Generally, this method can be used to analyze the impact of various factors on the results, or to solve the comprehensive evaluation problems that change with time. Its core is to establish the parent sequence that changes

ISSN 2616-5872 Vol.3, Issue 4: 5-10, DOI: 10.25236/AJEE.2021.030402

with time according to certain rules, take the change of each evaluation object with time as a sub sequence, and calculate the correlation between each sub sequence and the parent sequence, Draw a conclusion according to the correlation. In the process of system development, if the change trend of the two factors is consistent, that is, the degree of synchronous change is high, it can be said that the degree of correlation between the two is high; On the contrary, it is lower.

The specific calculation steps of grey system correlation analysis are as follows:

(1) Determine analysis sequence.

(2) Preprocessing variables.

(3) Calculate the correlation coefficient between each index in the subsequence and the reference sequence.

(4) Calculate the grey weighted correlation degree.

The grey correlation results of the sub index (IAQIP) of each pollutant project to the air quality index (AQI) are shown in Table 2:

Table 2: Grey correlation result analysis.

PM_{10}	NO_2	$PM_{2.5}$	O_3	CO	SO_2
0.979	0.974	0.96	0.95	0.933	0.915

According to the above table, the correlation (correlation degree) between six evaluation items (SO_2, CO, NO_2, PM_{10}, $PM_{2.5}$, O_3) and AQI is studied, and the analysis reference is provided based on the correlation degree. When using the grey correlation degree analysis, the resolution coefficient is 0.5, the correlation coefficient value is calculated in combination with the correlation coefficient calculation formula, and the correlation coefficient value is calculated according to the correlation coefficient value for evaluation and judgment. The resolution coefficient is defined $\rho \in (0,\infty)$, ρ the smaller, the greater the resolution, generally range of ρ is (0,1), and the specific value can be determined according to the situation $\rho \le 0.5463$, the resolution is the best, usually $\rho = 0.5$.

3.2. Spearman rank correlation coefficient method

In statistics, Pearson correlation coefficient is used to measure the correlation (linear correlation) between two variables X and Y, and its value is between -1 and 1. The Pearson correlation coefficient between two variables is defined as the quotient of the covariance difference and standard deviation between two variables.

The analysis results of Spearman rank correlation coefficient method are as follows:

(1) NO_2 concentration decreases with the increase of temperature, wind speed, humidity and wind direction, and increases with the increase of air pressure. This is because low temperature is not conducive to the photochemical reaction in the air, low wind speed is not conducive to the diffusion and transportation of pollutants, and high humidity makes it difficult for particles in the air to diffuse, thus affecting the refraction of light and hindering the photochemical reaction. The high pressure indicates that the meteorological conditions are stable, and the comprehensive influence leads to the change of NO_2 concentration in the air.

(2) CO concentration decreases with the increase of wind speed, humidity and temperature, and increases with the increase of wind direction and air pressure.

(3) O_3 concentration decreases with the increase of air pressure and humidity, and increases with the increase of wind speed, wind direction and temperature. This shows that there is a certain correlation between ambient air quality, ozone concentration and meteorological conditions, mainly with air temperature. The data show that there is still a certain lag between the daily maximum concentration of ozone and temperature, which mainly shows that the gradual formation of ozone under external factors takes a time process.

(4) $PM_{2.5}$ concentration and PM_{10} concentration decrease with the increase of wind direction, temperature, wind speed and humidity, and increase with the increase of air pressure. Strong pressure control, weak wind speed and high humidity are not conducive to the dilution, removal and diffusion of air pollutants, resulting in aggravation of pollution, and the pollution of $PM_{2.5}$ and PM_{10} seriously exceeds the standard

ISSN 2616-5872 Vol.3, Issue 4: 5-10, DOI: 10.25236/AJEE.2021.030402

4. Prediction of pollutant concentration based on WNN

Air quality prediction system is a multivariable, incident and complex nonlinear system. This complexity brings great difficulties to air quality prediction. In order to solve this problem, a secondary prediction system based on wavelet neural network is designed in this paper.

4.1. Wavelet neural network

Wavelet neural network replaces the node transfer function of neural network hidden layer by wavelet function. The weight from input layer to hidden layer and the threshold of hidden layer are replaced by scale scaling factor and time translation factor of wavelet function respectively.

In this paper, the secondary prediction model of pollutant is established based on wavelet neural network. The input layer is the primary prediction results, temperature, humidity, air pressure, wind speed and wind direction of the pollutant, and the output layer of the network is the measured data of the pollutant.

In this paper, Morlet mother wavelet basis function is used as the transfer function, as shown in formula:

$$\varphi(\mathrm{t}) = \mathrm{e}^{-\mathrm{t}^2/2}\mathrm{cosct} \tag{4}$$

The weight correction algorithm of wavelet neural network is similar to that of BP neural network. The correction process is as follows:

(1) Calculation prediction error:

$$\mathrm{e} = \frac{1}{2}\sum_{\mathrm{i}=1}^{\mathrm{L}}(\mathrm{y} - \hat{\mathrm{y}})^2 \tag{5}$$

(2) According to the error correction weight and coefficient:

$$\begin{cases} w_{n,k}^{i+1} = w_{n,k}^{i} + \Delta w_{n,k}^{i+1} \\ a_k^{i+1} = a_k^{i} + \Delta a_k^{i+1} \\ b_k^{i+1} = b_k^{i} + \Delta b_k^{i+1} \end{cases} \tag{6}$$

Where $\Delta w_{n,k}^{i+1}$, Δa_k^{i+1}, Δb_k^{i+1} is obtained from the following formula, where η is the learning rate.

$$\begin{cases} \Delta w_{n,k}^{i+1} = -\eta\Delta\frac{\partial e}{\partial w_{n,k}^{i}} \\ \Delta a_k^{i+1} = -\eta\Delta\frac{\partial e}{\partial a_k^{i}} \\ \Delta b_k^{i+1} = -\eta\Delta\frac{\partial e}{\partial b_k^{i}} \end{cases} \tag{7}$$

4.2. Wavelet network training steps

Step 1: data cleaning. Due to the duplication and missing parts of the original data, we first preprocess the data, fill the missing values with the method of filling in the mean value before and after the time, observe the data globally and locally in a visual way, and sort out 500-800 groups of effective data from the hourly prediction data and measured data of the monitoring points.

Step 2: network initialization. Set the number of network input nodes to 7 and the number of output nodes to 1. Random initialization weight and wavelet function coefficient scaling factor a_k and migration factor b_k. The learning rate is set to 0.001.

Step 3: train the network. The data is divided into training set and verification set, input into the network, calculate the error between the secondary prediction result and the measured value, and carry out backward propagation.

Step 4: end of training. End the training after 20000 iterations. According to the original data,

ISSN 2616-5872 Vol.3, Issue 4: 5-10, DOI: 10.25236/AJEE.2021.030402

the hourly pollutant concentration in the next three days is predicted.

4.3. Prediction results and error analysis

After the network training, draw and analyze the primary prediction and measured data of pollutants, the hourly secondary prediction results in the next three days, and the hourly prediction, secondary prediction and measured results of pollutants in the next three days.

Mean square error (MSE) is used to analyze the errors between the primary prediction results and the secondary prediction results and the measured data. The mean square error formula is:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{8}$$

Table 3: Primary and secondary prediction MSE of pollutants at monitoring point A.

Types of pollutants	SO_2	NO_2	PM_{10}	$PM_{2.5}$	CO	O_3
Primary forecast MSE	1.1441	4.6659	3.0270	1.8388	0.0421	5.2222
Quadratic prediction MSE	0.8155	1.1812	4.9134	2.7278	0.0555	1.8779

For the prediction of pollutants at the monitoring points, the secondary prediction accuracy of SO_2, NO_2 and O_3 has been significantly improved, but the secondary prediction effect of PM10 and PM2.5 is not ideal.

From 0:00 to 7:00 on July 13, 2021, the comparison of hourly primary prediction, secondary prediction and measured results of pollutants is shown in the figure below.

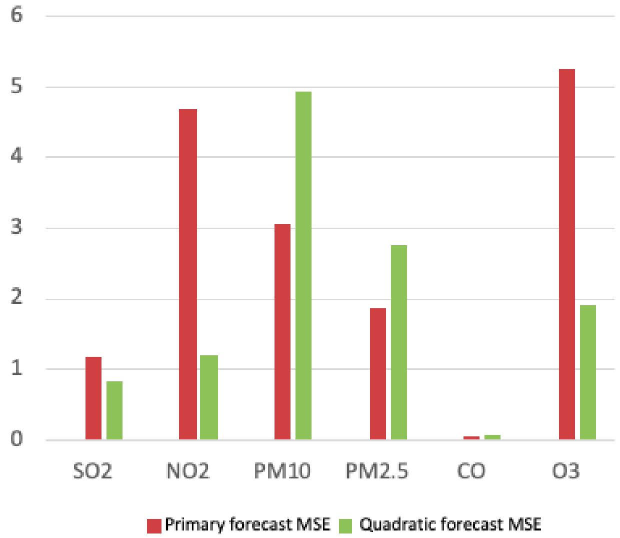

Figure 1: Comparison of MSE between primary prediction and secondary prediction at monitoring points.

Overall, the secondary prediction model of pollutant concentration based on wavelet neural network improves the accuracy of the primary prediction results, especially for SO_2 and NO_2, but the prediction effect for O_3 is not ideal. Due to the lack of original training data, the prediction error of the trained neural network model for some pollutant concentrations is too large. When there are enough original data, the prediction of air quality index by WNN neural network model will have higher accuracy, and its calculation efficiency will be improved in the long run. In the previous paper, we found that there is a certain correlation between pollutant concentration and meteorological factors, so we can consider meteorological factors as part of the input of the prediction model to improve the accuracy of the model.

5. Conclusion

In this paper, the change of air pollutant concentration is studied and predicted based on gray

ISSN 2616-5872 Vol.3, Issue 4: 5-10, DOI: 10.25236/AJEE.2021.030402

correlation analysis, the characteristics of various meteorological conditions are analyzed, and the impact of meteorological conditions on air pollutant concentration is studied. The prediction model based on WNN improves the prediction accuracy based on the primary prediction model, especially for the NO_2 concentration prediction accuracy of monitoring point a, the mean square error is reduced by 75%. Considering that the pollutant concentrations in adjacent areas often have a certain correlation, regional collaborative prediction may improve the accuracy of air quality prediction. Therefore, on the basis of this prediction model, increasing the influencing factors of adjacent areas can be used as the direction of future realization.

References

[1] Hao Jiming, Ma Guangda, Wang Shuxiao. Air pollution control engineering [M]. Beijing: Higher Education Press, 2010.

[2] Liu Qing, Li Dian, Wang Liwei, Xu Yaqi, Wu Yutong. Study on correlation modeling between air pollutant concentration variation characteristics and meteorological factors [J]. Environmental science and management, 2021, 46 (04): 136-140.

[3] Chang Ling, Zhang Feng, Liu Dong, Wu Xiaoyan, Ma Jinlong. Analysis of indoor air pollution status and related factors in Xi'an [J]. Public health and preventive medicine, 2021, 32 (04): 45-49.

[4] Li Xiaolin, Yang Xi. Analysis on the characteristics of ambient air quality and its correlation with meteorological factors in Zigong City from 2016 to 2018. Sichuan environment, 2020, 39 (03): 59-66.

[5] Zhao Zijing, Wei Yongjie, Zhang Xiangzhi, Qin Wei, Xie Huifang. Correlation analysis between haze weather and main meteorological conditions in Nanjing [J]. China Environmental Science, 2015, 35 (12): 3570-3580.