Establishment and Effect Evaluation of Prediction Models of Ozone Concentration in Baoding City

Xiangru KONG[1], Jiajia ZHANG[1], Luntao YAO[1], Tianning YANG[2], Rongfang YANG[1,3,4]*

1. Hebei Meteorological Technology and Equipment Center, Shijiazhuang 050021, China; 2. Weichang Meteorological Bureau, Weichang 068450, China; 3. Xiong'an Key Laboratory of Atmospheric Boundary Layer, China Meteorological Administration, Xiong'an New Area 071800, China; 4. Hebei Key Laboratory of Meteorology and Ecological Environment, Shijiazhuang 050021, China

Abstract Firstly, based on the data of air quality and the meteorological data in Baoding City from 2017 to 2021, the correlations of meteorological elements and pollutants with O_3 concentration were explored to determine the forecast factors of forecast models. Secondly, the O_3-8h concentration in Baoding City in 2021 was predicted based on the constructed models of multiple linear regression (MLR), backward propagation neural network (BPNN), and auto regressive integrated moving average (ARIMA), and the predicted values were compared with the observed values to test their prediction effects. The results show that overall, the MLR, BPNN and ARIMA models were able to forecast the changing trend of O_3-8h concentration in Baoding in 2021, but the BPNN model gave better forecast results than the ARIMA and MLR models, especially for the prediction of the high values of O_3-8h concentration, and the correlation coefficients between the predicted values and the observed values were all higher than 0.9 during June – September. The mean error (ME), mean absolute error (MAE), and root mean square error (RMSE) of the predicted values and the observed values of daily O_3-8h concentration based on the BPNN model were 0.45, 19.11 and 24.41 μg/m^3, respectively, which were significantly better than those of the MLR and ARIMA models. The prediction effects of the MLR, BPNN and ARIMA models were the best at the pollution level, followed by the excellent level, and it was the worst at the good level. In comparison, the prediction effect of BPNN model was better than that of the MLR and ARIMA models as a whole, especially for the pollution and excellent levels. The TS scores of the BPNN model were all above 66%, and the PC values were above 86%. The BPNN model can forecast the changing trend of O_3 concentration more accurately, and has a good practical application value, but at the same time, the predicted high values of O_3 concentration should be appropriately increased according to error characteristics of the model.

Key words Ozone (O_3); Multiple linear regression model; Back propagation neural network model; Auto regressive integrated moving average model; TS

DOI 10.19547/j.issn2152 – 3940.2025.03.010

With the acceleration of industrialization and the rapid increase in the number of motor vehicles in China, air pollution in China is transforming from a single type of air pollution to a complex type of air pollution[1]. During the 13th Five-Year Plan period, particulate matter pollution in China has been effectively prevented and controlled since the implementation of strict measures for the control of air pollution. However, O_3 pollution has shown a significant upward trend and has become the main factor restricting the improvement of air quality in China at the current stage. High concentrations of O_3 can not only pose a threat to the physical and mental health of residents[2–4], but also have a significant impact on climate change, ecological environment, *etc*[5–7]. Therefore, accurately forecasting the status of O_3 pollution in cities is crucial for establishing an effective mechanism for the early warning of air pollution and safeguarding public health[8–11].

Received: April 16, 2025 Accepted: May 20, 2025

Supported by the Project of the Key Open Laboratory of Atmospheric Detection, China Meteorological Administration (2023KLAS02M); the Second Batch of Science and Technology Project of China Meteorological Administration ("Jiebangguashuai"): the Research and Development of Short-term and Near-term Warning Products for Severe Convective Weather in Beijing – Tianjin – Hebei Region (CMAJBGS202307).

* Corresponding author.

The domestic forecast of O_3 concentration is mainly divided into numerical model forecast and statistical model forecast. Numerical model forecast mainly relies on meteorological data and emission inventories of pollution sources to simulate air quality on a large scale in both horizontal and vertical directions. Due to the difficulty in establishing localized inventories, the complexity of model operation, and huge calculation amount, numerical forecasting methods are difficult to be widely applied[12–13]. Statistical model forecast means that based on pollutant concentration and meteorological observation data, statistical methods such as multiple linear regression, neural networks, and time series are used to establish numerical relationship models between pollutants and meteorological factors, and it has the advantages of easy data acquisition, small computational load, and fast operation speed[14–15], so it has been widely applied in the forecast of regional air quality[16–18]. Chen Chen *et al.*[19] established an equation for the forecast of O_3 concentration in Foshan area in 2018 using the multiple linear regression method, and found that the correlation coefficient between predicted and observed values could reach 0.82, indicating that the forecast effect of the multiple linear regression equation was good. Based on the observation data of O_3 concentration from Tianjin, Shangdianzi and Baoding stations, Li Yingruo *et al.*[20] conducted the medium- and long-term forecast of O_3 concentration using the ARIMA model. The results show that the correlation coefficient between the monthly average of predicted and observed values could reach 0.951, and the root mean square error was only 10.2 $\mu g/m^3$, indicating a relatively ideal forecast effect. Zhang Wenfang[21] constructed a BPNN forecast model using the data of meteorological factors and O_3 concentration from six monitoring points in the northwest of Beijing urban area, and then verified the forecast results by combining the observed values of O_3 concentration. The results reveal that the correlation coefficient between the predicted and observed values of O_3 concentration was as high as 0.883, so the forecast accuracy was relatively high. Du Qinbo *et al.* [22] used BPNN and stepwise linear regression models to predict O_3 concentration in Shantou City from 2014 to 2017, and found that both BPNN and linear regression models could reflect the changing trend of O_3 concentration, but the forecast result of BPNN model was better than that of the linear regression model. Based on the reanalysis data of O_3 concentration and ERA5 in Haikou City from 2015 to 2020, Fu Chuanbo *et al.* [23] established three prediction models for O_3 concentration: MLR, BPNN, and SVM (support vector machine), and predicted O_3-8h concentration in Haikou City in 2021. The results indicate that the prediction effect of BPNN was superior to that of the other two prediction models.

Baoding City, which is located in the Beijing – Tianjin – Shijiazhuang region, is one of the central cities in the Beijing – Tianjin – Hebei region approved by the State Council and also an important city that takes the lead in coordinated development in the Beijing – Tianjin – Hebei region. In terms of terrain, it is high in the northwest and lower in the southeast. Due to the barrier effect of the Taihang Mountains, the annual average wind speed is relatively low, and there are frequent still winds and light winds. The conditions for pollutant diffusion are poor, and the problem of O_3 pollution is relatively prominent due to the accumulation of external transmission and local source emissions[24–25]. According to the monitoring data of environmental air quality in Baoding City in 2021, the proportion of days with O_3 as the primary pollutant in the total polluted days in Baoding City in 2021 was as high as 43.6%, so the problem of ozone pollution cannot be ignored. Therefore, to solve the current problem of ozone pollution in Baoding City, based on the monitoring data of air quality in Baoding City and the meteorological observation data of Baoding National Basic Meteorological Station from 2017 to 2020, the MLR, BPNN and ARIMA models for the prediction of O_3 concentration in Baoding City were constructed, and the predicted values of O_3 concentration in 2021 were compared with the observed values, so as to provide a basis and reference for formulating effective strategies for the prevention and control of O_3 pollution.

1 Data and methods

1.1 Data sources The data of ambient air quality were from China National Environmental Monitoring Centre (http://www.cnemc.cn/sssj/). The hourly monitoring data of PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO and O_3 from nine national control monitoring stations for atmospheric environmental quality in Baoding City during 2017 – 2021 were selected for research. The meteorological data were derived from the concurrent observation data of Baoding National Basic Meteorological Station (station number: 54602). The distribution of the national control monitoring stations for ambient air quality and the meteorological station is shown in Fig. 1.

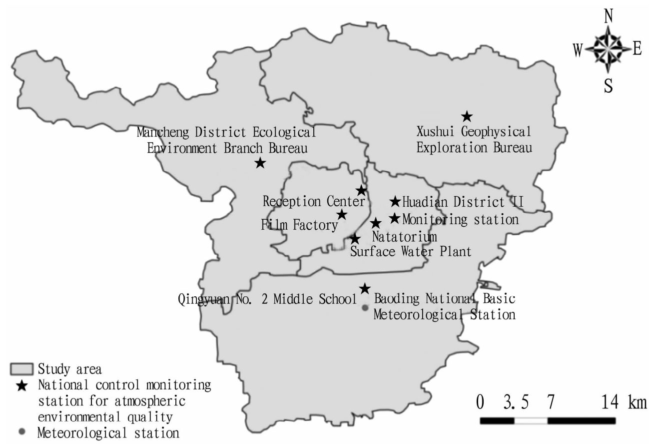


Fig. 1 Distribution of the national control monitoring stations for atmospheric environmental quality and the meteorological station in Baoding City

1.2 Research methods

1.2.1 Statistical methods. According to the *Ambient Air Quality Standards* (GB 3095 – 2012), O_3-8h (short for the daily maximum 8-hour moving average of O_3 concentration) was selected as the evaluation indicator for daily O_3 concentration, and its secondary concentration limit is 160 $\mu g/m^3$. According to the *Technical Regulation on Ambient Air Quality Index* (HJ 633 – 2012), $\rho(O_3$-8h) can be divided into three grades: 0 – 100 $\mu g/m^3$ (excellent), 101 – 160 $\mu g/m^3$ (good), and >160 $\mu g/m^3$ (O_3 pollution).

The individual index and composite index can be calculated in accordance with the *Technical Regulation on the Ranking of Urban Environmental Air Quality*[26], and the formulas are as follows:

$$I_i = \frac{C_i}{S_i}$$

$$I_{sum} = \sum_{i=1}^{6} I_i$$

In the formulas, S_i represents the standard value of indicator i, corresponding respectively to the secondary standard limits of the annual average concentrations of PM_{10}, $PM_{2.5}$, SO_2 and NO_2, the secondary standard limit of the daily average concentration of CO, and the secondary standard limit of O_3-8h concentration; C_i means the evaluation concentration of indicator i. In this paper, C_i is the annual average concentration of PM_{10}, $PM_{2.5}$, SO_2 and NO_2, as well as the annual average concentration based on the 95^{th} percentile of CO and the 90^{th} percentile of O_3-8h.

The contribution rate of the composite index of pollutants (U_i) is the ratio of the index of a single pollutant to the composite index of six pollutants[27], and the formula is as follows:

$$U_i = \frac{I_i}{I_{sum}} \times 100\%$$

1.2.2 Modeling methods.

1.2.2.1 Multivariable linear regression (MLR) model[28]. A MLR model refers to a linear regression model containing multiple explanatory variables, and is used to explain the linear relationship between the explained variable and many other explanatory variables. Its mathematical model is as follows:

$$y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_P X_P + \varepsilon$$

Here, y refers to the dependent variable; P is the number of explanatory variables; X is the explanatory variable; β is the coefficient of the explanatory variable; ε is the random error.

1.2.2.2 Auto regressive integrated moving average (ARIMA) model[29]. ARIMA is one of the methods for the predictive analysis of time series. ARIMA (p, d, q) combines the auto regressive model, moving average model and difference method. Its formula is defined as follows:

$$Y_i = \mu + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \varepsilon_i + \sum_{i=0}^{q} \beta_i \varepsilon_{t-i}$$

In the formula, Y_i is the current value; μ is the constant term; ε_i is the error; p and q are the orders of the auto regressive model and moving average model, respectively; α_i and β_i are the correlation coefficients of the two models, respectively; d is the number of differences (order) made to make them stationary sequences.

1.2.2.3 Back propagation neural network (BPNN) model[30]. BPNN means that the gradient of each parameter in the neural network to the objective function is calculated by using the chain rule[31], and then the parameters are continuously updated by using optimization algorithms such as gradient descent until the prediction accuracy of the model is satisfactory.

1.2.3 Analytical method. The predicted and observed values were statistically analyzed by using error indicators and graded evaluation method[32]. Eight indicators, namely mean absolute error (MAE), mean error (ME), root mean square error (RMSE), correlation coefficient (R), accuracy rate (PC), threat score (TS), false alarm rate (FAR), and missing report rate (PO), were selected for evaluation[33]. The formula of each indicator is shown as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |(P_i - O_i)|$$

$$ME = \frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (P_i - O_i)^2}$$

$$R = \frac{\sum_{i=1}^{N} (P_i - \overline{P})(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{N} (P_i - \overline{P})^2} \sqrt{\sum_{i=1}^{N} (O_i - \overline{O})^2}}$$

$$PC = \frac{NA + ND}{NA + NB + NC + ND} \times 100\%$$

$$TS = \frac{NA}{NA + NB + NC} \times 100\%$$

$$FAR = \frac{NB}{NA + NB} \times 100\%$$

$$PO = \frac{NC}{NA + NC} \times 100\%$$

In the formulas, P_i represents the predicted value; O_i stands for the observed value; N means the number of samples. $\overline{P}$ is the average of the predicted values; $\overline{O}$ represents the average of the observed values. NA means the number of days when both the predicted and observed values occur; NB represents the number of days when the predicted values occur but the observed values do not appear; NC is the number of days when the predicted values do not occur but the observed values appear; ND represents the number of days when neither the predicted values nor the observed values occur.

Among them, the closer the absolute value of R is to 1, the higher the linear correlation between the two[34]. The closer the MAE, ME and RMSE are to 0, the better the prediction results will be. The TS and PC values range from 0 to 100%, and the closer the values are to 100%, the better the prediction effect[35]. The FAR and PO values range from 0 to 100%, and the lower the values, the higher the prediction accuracy.

2 Analysis of the impact of O_3 on air quality in Baoding City

Fig. 2a shows the annual variation of the proportion of days with O_3 as the primary pollutant during 2017 – 2021. It can be seen from the figure that the proportion of days with O_3 as the primary pollutant in Baoding City showed a fluctuating upward trend from 2017 to 2021. In 2019, the proportion reached 42.7%, 8.9% higher than that in 2017. In 2021, it was up to 40.7%. Fig. 2b presents the contribution of O_3 to the composite index of ambient air quality during 2017 – 2021. As shown in the figure, the contribution of O_3 to the composite index of air quality increased year by year from 2017 to 2021. In 2021, the contribution rate was the highest, reaching 21.9%, increasing by 5.6% compared with 2017. In conclusion, the situation of O_3 pollution in Baoding City was significantly aggravated during 2017 – 2021, and O_3 had become the primary pollutant affecting the ambient air quality of the city.

3 Test on the forecast effect of different statistical forecast models on O_3-8h concentration

3.1 Selection of forecast factors Previous studies reveal that O_3 concentration is jointly affected by meteorological conditions and pollutants[36-39]. Table 1 presents the correlation of O_3-8h concentration with temperature (T), relative humidity (RH), air pressure (P), 10-meter wind speed (W_{10}), wind direction (F), as well as SO_2, NO_2, CO, $PM_{2.5}$, and PM_{10}. It can be seen from the table that the correlation of O_3-8h concentration with SO_2, NO_2, CO, $PM_{2.5}$, T, RH, P and W_{10} all passed the significance test with $\alpha = 0.01$, and their correlation coefficients were all above 0.25. Among them, T had the best correlation with O_3-8h concentration, and the correlation coefficient could reach 0.80. Therefore, SO_2, NO_2, CO, $PM_{2.5}$, T, RH, P and W_{10} were selected as the prediction factors of O_3-8h concentration.

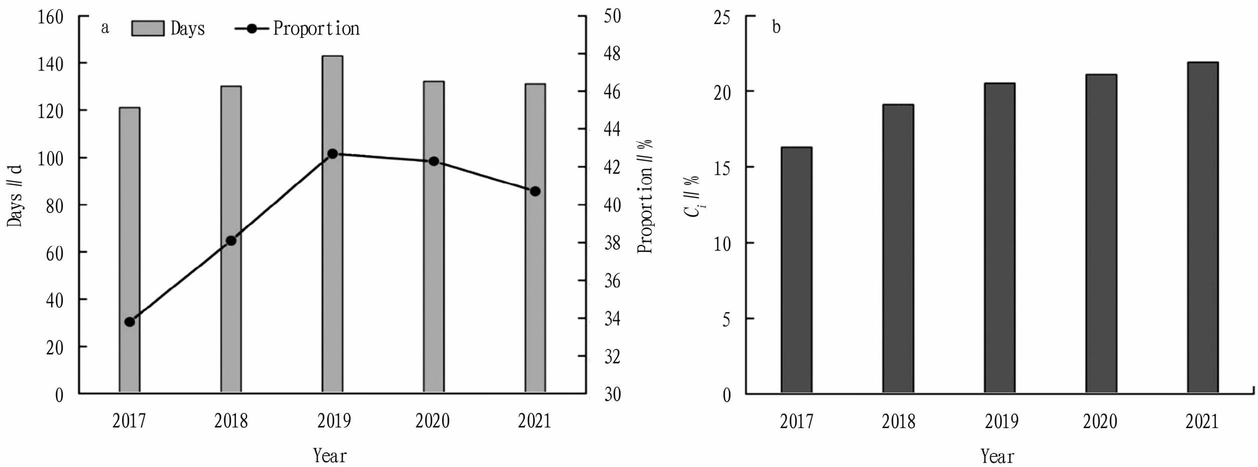


Fig. 2 Proportion of days with O_3 as the primary pollutant (a) and the contribution rate of O_3 to the composite index of air quality (b) in Baoding City during 2017 – 2021

Table 1 Correlation of O_3-8h concentration with meteorological factors and pollutants in Baoding City

R	SO_2	NO_2	CO	$PM_{2.5}$	PM_{10}	T	RH	P	W_{10}	F
O_3	–0.25**	–0.68**	–0.47**	–0.32**	–0.09*	0.80**	–0.28**	–0.60**	–0.28**	0.10*

Note: ** means passing the significance test with $\alpha = 0.01$, and * means passing the significance test with $\alpha = 0.05$.

3.2 Construction of forecast models Based on the daily average concentrations of $\rho(SO_2)$, $\rho(NO_2)$, $\rho(CO)$, $\rho(PM_{2.5})$, $\rho(PM_{10})$, RH, W_{10}, T, and P in Baoding City from 2017 to 2020, the MLR, BPNN, and ARIMA forecast models of O_3-8h concentration in Baoding City were respectively constructed as follows.

(1) The MLR forecast model of O_3-8h concentration in Baoding City was constructed by using SPSS software:

$$\rho(O_3\text{-8h}) = 0.391 \times \rho(SO_2) - 0.741 \times \rho(NO_2) + 2.245 \times \rho(CO) - 0.531 \times \rho(PM_{2.5}) + 0.629 \times \rho(PM_{10}) + 3.572 \times T - 0.613 \times RH + 0.008 \times P - 8.012 \times W_{10} + 90.241$$

In the formula, ρ represents the regression value obtained by the multiple linear regression, and this equation passed the significance test with $\alpha = 0.01$.

(2) The BPNN model was constructed based on the neural network module of SPSS software. 70% of the data were randomly selected as the training dataset, and the remaining 30% of the data were used as the validation dataset. The weight of each parameter of the BPNN model was debugged. After the model was stable, the forecast factors in 2021 were substituted into the model for forecast.

(3) The data of $\rho(SO_2)$, $\rho(NO_2)$, $\rho(CO)$, $\rho(PM_{2.5})$, $\rho(PM_{10})$, RH, W_{10}, T, P and O_3-8h concentration from 2017 to 2020 were preprocessed, and the ARIMA model was constructed by using the time series prediction module of SPSS software. Finally, the forecast results of O_3-8h concentration in 2021 were obtained.

3.3 Test of forecast effect

3.3.1 Monthly variation. The monthly averages of predicted and observed values of O_3-8h concentration in Baoding City in 2021 are shown in Fig. 3, and their correlation coefficients are shown in Table 2. As shown in Fig. 3 and Table 2, the MLR, BPNN and ARIMA models could all predict the monthly variation characteristics of O_3-8h concentration, which is similar to the research results of Wang Xinlu *et al.* [40] and Fu Chuanbo *et al.* [23]. In comparison, the correlation between the observed values and predicted values by the BPNN model was the best, and the correlation coefficients were all above 0.22. Particularly, its simulation effect on high values of O_3-8h concentration was superior to that of the MLR and ARIMA models. The correlation coefficients were all above 0.9 from June to September, up to 0.93 in July, significantly higher than that of the MLR and ARIMA models. However, the model underestimated the high values of O_3-8h concentration. The correlation between the observed values and predicted values by the MLR model was relatively poor, and the model overestimated the high values of O_3-8h concentration, but its prediction effect on low values of O_3-8h concentration was better. The correlation between the observed values and predicted values by the ARIMA model was the poorest, and its prediction effect on high values of O_3-8h concentration was also poor, but its prediction effect on low values of O_3-8h concentration was better.

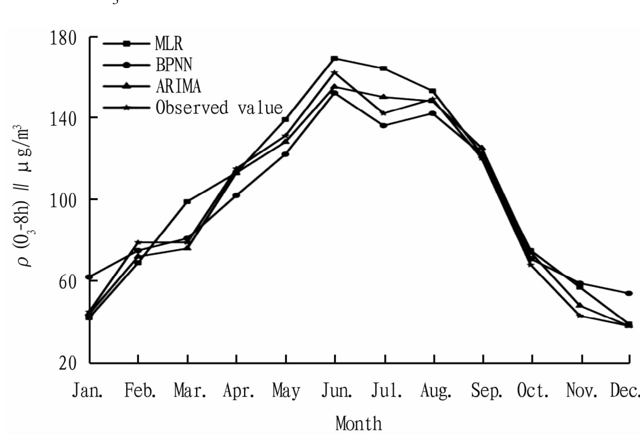


Fig. 3 Comparison between the observed and predicted values of O_3-8h concentration based on the three statistical models in Baoding City in 2021

Table 2 Statistical test on the prediction results of O_3-8h concentration based on the three models

Model	January	February	March	April	May	June	July	August	September	October	November	December
MLR	0.25**	0.49**	0.10	0.05	0.57**	0.89**	0.87**	0.82**	0.90**	0.44**	0.66**	0.53**
BPNN	0.47**	0.26**	0.43**	0.55**	0.84**	0.91**	0.93**	0.92**	0.91**	0.27**	0.71**	0.62**
ARIMA	0.18*	0.50**	0.40**	0.32**	0.40**	0.26**	0.66**	0.35**	0.49**	0.22**	0.55**	0.24**

Note: ** means passing the significance test with $\alpha = 0.01$, and * means passing the significance test with $\alpha = 0.05$.

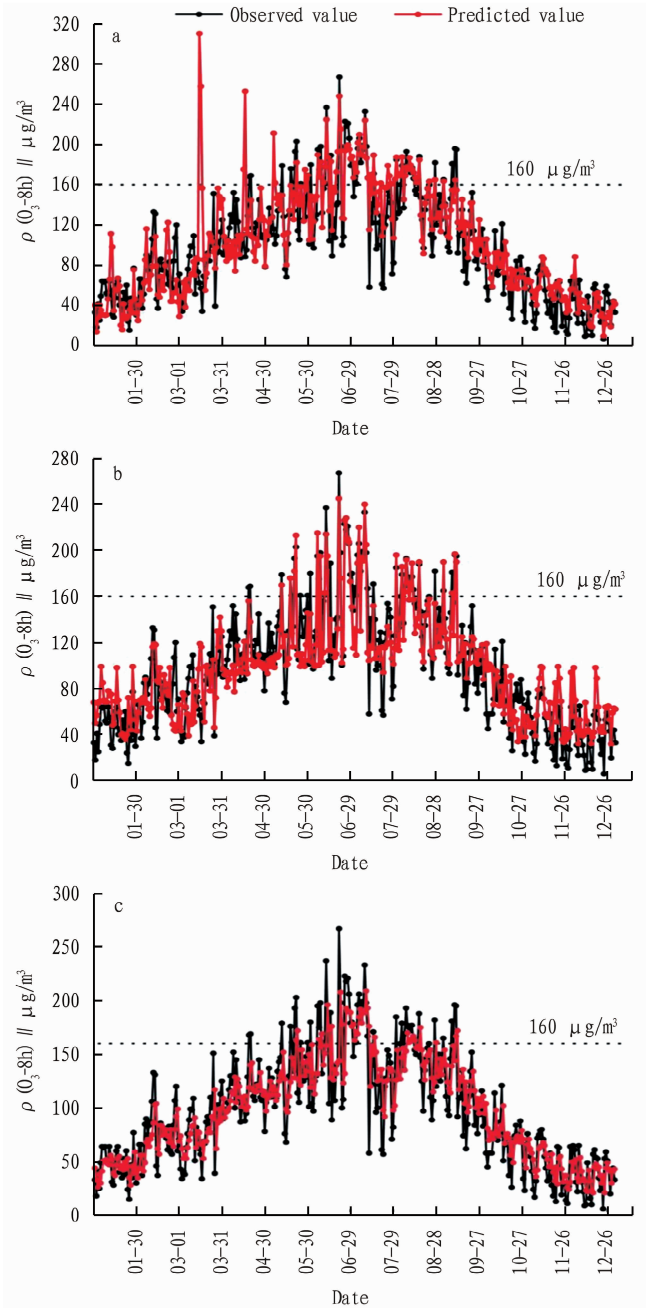
3.3.2 Daily variation. Fig. 4 shows the daily averages of predicted and observed values of O_3-8h concentration in Baoding City in 2021, and their correlation coefficients are shown in Table 3. From Fig. 4 and Table 3, the correlation coefficients between the observed values and predicted values of O_3-8h concentration by the MLR, BPNN and ARIMA models were 0.84, 0.88 and 0.85, respectively, all passing the significance test with $\alpha = 0.01$. That is, all three models could well simulate the daily variation trend of O_3-8h concentration. The ME values of the ARIMA and BPNN models were -0.01 and 0.45 $\mu g/m^3$, respectively, that is, the deviation degree between the predicted values of the ARIMA and BPNN models and the observed values was relatively small. The ME value of the MLR model was up to 6.12 $\mu g/m^3$, that is, the deviation degree between the predicted values of the MLR model and the observed values was relatively large, which is mainly related to the poor prediction results of the MLR model for O_3-8h concentration in the spring of 2021. The MAE and RMSE values of the BPNN model were 19.11 and 24.41 $\mu g/m^3$, respectively, significantly smaller than those of the MLR and ARIMA models, indicating that the prediction effect of the BPNN model was the most ideal. On the whole, it can be concluded that the MLR, BPNN and ARIMA models could all predict the changing trends of O_3-8h concentration relatively well. Relatively speaking, the prediction effect of the BPNN model was better than that of the MLR and ARIMA models.

Table 3 Statistical test on the prediction results of O_3-8h concentration based on the three models

Model	R	R^2	MAE	ME	RMSE
MLR	0.84**	0.70**	20.04	6.12	30.05
BPNN	0.88**	0.78**	19.11	0.45	24.41
ARIMA	0.85**	0.71**	21.34	-0.01	28.09

Note: ** means passing the significance test with $\alpha = 0.01$, and * means passing the significance test with $\alpha = 0.05$.

3.4 Test on the accuracy of forecast grades Four indicators, namely TS, PC, PO, and FAR, were used to further explore the accuracy of O_3-8h concentration predicted by the MLR, BPNN and ARIMA models at different levels of O_3-8h concentration. It can be seen from Table 4 that as the level of O_3-8h concentration in Baoding City was excellent, the TS scores of the MLR, BPNN and ARIMA models were 76.61%, 77.43% and 78.70%, respectively, and the PC values were 86.03%, 86.03% and 87.40%, respectively. The TS and PC values of the three models were all greater than 75%, and the FAR and PO values ranged from 8.08% to 13.53%, indicating that the prediction accuracy of the three models was relatively high when the level of O_3-8h concentration was excellent, and there were fewer false reports and missed reports. When the level of O_3-8h concentration in Baoding City was good, the TS scores of the MLR, BPNN and ARIMA models



Note: a. MLR; b. BPNN; c. ARIMA.

Fig. 4 Comparison between the observed and predicted values of O_3-8h concentration based on the three statistical models in Baoding City in 2021

were 50.32%, 59.24% and 53.29%, and the PC values were 78.90%, 82.47% and 78.63%, respectively, which were significantly lower than those of the excellent grade. The FAR and PO values ranged from 8.24% to 18.12%, both of which were higher than those of the excellent grade. For days with O_3 pollution, the TS scores of the MLR, BPNN and ARIMA models were 53.95%, 66.67% and 39.68%, respectively, slightly lower than those of the good grade. However, their PC values were 90.41%, 95.34% and 89.59%, respectively, significantly higher than those of the excellent and good grades. Moreover, both their FAR values and PO values were lower than 7.72%, significantly lower than those of the excellent and good grades. Overall, the prediction effects of the MLR, BPNN and ARIMA models were the most ideal at the pollution level, followed by the excellent level, and it was the worst at the good level.

The prediction effects of the MLR, BPNN and ARIMA models at different grades of air quality were compared. Generally speaking, the TS and PC values of the BPNN model at different grades of air quality were higher than those of the MLR and ARIMA models, revealing that the prediction accuracy of the BPNN model at different grades of air quality was superior to that of the MLR and ARIMA models, which is similar to the research result of Fu Chuanbo *et al*[23].

Table 4 Accuracy of O_3-8h concentration predicted by the three models at different levels of O_3-8h concentration in 2021 %

Model	Excellent				Good				Pollution			
	TS	PC	FAR	PO	TS	PC	FAR	PO	TS	PC	FAR	PO
MLR	76.61	86.03	8.08	13.53	50.32	78.90	13.24	13.03	53.95	90.41	7.72	2.83
BPNN	77.43	86.03	12.63	11.09	59.24	82.47	14.71	8.24	66.67	95.34	0	4.27
ARIMA	78.70	87.40	9.74	10.60	53.29	78.63	18.12	10.03	39.68	89.59	3.53	7.07

4 Conclusions

Based on the data of environmental air quality in the urban area of Baoding City from 2017 to 2021, as well as the meteorological observation data of the same period, the MLR, ARIMA and BPNN models for the forecast of O_3-8h concentration in Baoding City was constructed, and the predicted values and observed values of O_3-8h concentration in Baoding City in 2021 were compared and tested. The main conclusions are as follows.

(1) The MLR, BPNN and ARIMA models could all predict the monthly variation trends of O_3-8h concentration. The correlation between the observed values and predicted values by the BPNN model was the best. Particularly, its simulation effect on high values of O_3-8h concentration was superior to that of the MLR and ARIMA models. The correlation coefficients were all above 0.9 from June to September. From the perspective of daily variation, the correlation coefficient between the observed values and the predicted values of the BPNN model was as high as 0.88, and its ME, MAE and RMSE values were 0.45, 19.11 and 24.41 $\mu g/m^3$, respectively, which were significantly better than those of the MLR and ARIMA models. Overall, all the three models could predict the changing trends of O_3-8h concentration relatively well. Relatively speaking, the prediction effect of the BPNN model was superior to that of the MLR and ARIMA models.

(2) The prediction effects of the MLR, BPNN and ARIMA models at different levels of O_3-8h concentration show that the prediction effects of the three models were the best at the pollution level, followed by the excellent level, and it was the worst at the good level. In comparison, the prediction effect of the BPNN model was generally superior to that of the traditional MLR and ARIMA models, especially for the pollution and excellent levels. The TS scores of the BPNN model were above 66%, and the PC values were above 86%.

(3) In practical application, a BPNN prediction model can provide relatively accurate proximity forecasting results for air quality assurance during a major event, and has a good application value. However, it is also necessary to appropriately increase the predicted high values of O_3 concentration according to error characteristics of the model.

References

[1] HE WN, LI X, XIE SQ, *et al*. Characteristics $PM_{2.5}$ and O_3 pollution and their interaction in Taizhou urban area[J]. Environmental Monitoring in China, 2023, 39(1): 60–68.

[2] TANG XY, ZHANG YH, SHAO M. Atmospheric environmental chemistry[M]. Beijing: Higher Education Press, 2006: 268–327.

[3] YANG C, YANG H, GUO S, *et al*. Alternative ozone metrics and daily mortality in Suzhou: The China air pollution and health effects study (CAPES)[J]. Science of the Total Environment, 2012, 426(2): 83–89.

[4] WANG X, LU W, WANG W, *et al*. A study of ozone variation trend within area of affecting human health in Hong Kong[J]. Chemosphere, 2003, 52(9): 1405–1410.

[5] LI P, DE MARCO A, FENG ZZ, *et al*. Nationwide ground-level ozone measurements in China suggest serious risks to forests[J]. Environmental Pollution, 2018, 237: 803–813.

[6] FENG ZZ, DE MARCO A, ANAV A, *et al*. Economic losses due to ozone impacts on human health, forest productivity and crop yield across China[J]. Environment International, 131: 3–5.

[7] AGATHOKLEOUS E, FENG ZZ, OKSANEN E, *et al*. Ozone affects plant, insect, and soil microbial communities: A threat to terrestrial ecosystems and biodiversity[J]. Science Advances, 2020, 6(33): 4–5.

[8] WANG P, CHEN Y, HU J, *et al*. Source apportionment of summertime ozone in China using a source-oriented chemical transport model[J]. Atmospheric environment, 2019, 211: 79–90.

[9] HAKIM ZQ, ARCHER-NICHOLLS S, BEIG G, *et al*. Evaluation of tropospheric ozone and ozone precursors in simulations from the HTAPII and CCMI model intercomparisons: A focus on the Indian subcontinent[J]. Atmospheric Chemistry and Physics, 2019, 19(9): 6437–6458.

[10] XUE T, WANG R, TONG M, *et al*. Estimating the exposure-response function between long-term ozone exposure and under-5 mortality in 55

low-income and middle-income countries: A retrospective, multicentre, epidemiological study[J]. The Lancet Planetary Health, 2023, 7(9): 736–746.

[11] HU WY, ZHAO TL, ZHENG XB, *et al.* Comparative analysis of near-surface ozone concentration in typical large cities in central and eastern China[J]. Environmental Science & Technology, 2019, 42(S2): 173–179.

[12] YANG D. Ambient air quality forecast and early warning system model, uncertain factors and countermeasures [J]. Environment and Development, 2020, 32(10): 225–226.

[13] DING S, CHEN BZ, WANG J, *et al.* An applied research of decision-tree based statistical model in forecasting the spatial-temporal distribution of O_3[J]. Acta Scientiae Circumstantiae, 2018, 38(8): 3229–3242.

[14] CHEN YG. Prediction algorithm of $PM_{2.5}$ mass concentration based on adaptive BP neural network[J]. Computing Archives for Informatics & Numerical Computation, 2018, 100(8): 825.

[15] BIANCOFIORE F, BUSILACCHIO M, VERDECCHIA M, *et al.* Recursive neural network model for analysis and forecast of PM_{10} and $PM_{2.5}$ [J]. Atmospheric Pollution Research, 2017, 8(4): 652–659.

[16] SHEN J, ZHONG LJ, HE FF, *et al.* Development of air quality forecast model based on clustering and multiple regression[J]. Environmental Science & Technology, 2015, 38(2): 63–66.

[17] LU Q, WANG GH, FENG YC, *et al.* The influence of meteorological conditions on a heavy ozone pollution process in Chengde City[J]. Journal of Ecology and Rural Environment, 2019, 35(8): 992–999.

[18] LI ZM, ZHAO XJ, SUN ZB, *et al.* Research on the interpretation and correction of numerical ozone forecast based on analog ensemble[J]. China Environmental Science, 2020, 40(2): 475–484.

[19] CHEN C, HONG YY, TAN HB, *et al.* Establishment and application of Foshan ozone concentration forecast equation[J]. Environmental Science, 2022, 43(10): 4316–4326.

[20] LI YR, HAN TT, WANG JX, *et al.* Application of ARIMA model for mid- and long-term forecasting of ozone concentration[J]. Environmental Science, 2021, 42(7): 3118–3126.

[21] ZHANG WF. Study on the variation and prediction method of ozone concentration in Beijing urban area[D]. Beijing: Beijing Forestry University, 2010: 39–52.

[22] DU QB, CHEN HH, LI YY, *et al.* Application of BP neural network model in O_3 mass concentration forecasting[J]. Guangdong Meteorology, China, 2019, 41(3): 29–32.

[23] FU CB, LIN JX, TANG JX, *et al.* Establishment and effective evaluation of Haikou ozone concentration statistical prediction model[J]. Environmental Science, 2024(5): 2516–2524.

[24] KONG XR. Simulation study on the effect of ozone pollution improvement by emission reduction scenario of ozone precursors in the central urban area of Baoding[D]. Lanzhou: Lanzhou University, 2023: 1–4.

[25] ZHANG XY, WANG XQ, WANG CD, *et al.* Ozone sensitivity and precursor emission reduction scheme in Baoding City in summer[J]. China Environmental Science, 2023, 43(6): 1–15.

[26] DAI J, YANG J, CHEN MF, *et al.* Analysis of air pollution characteristics in Panzhihua City based on different time scale of the composite index contribution rate[J]. Environmental Monitoring in China, 2023, 39(S1): 52–58.

[27] LI QF, LU N, CHENG LP, *et al.* Analysis of air pollution characteristics in Shijiazhuang City based on pollution contribution rate[J]. Hebei Journal of Industrial Science and Technology, 2019, 36(1): 59–65.

[28] TU XQ. Prediction of water environment quality in Yongan small watershed based on neural network model[J]. Leather Manufacture and Environmental Technology, 2023, 4(20): 75–78.

[29] XU S, LIU DD. Power load forecasting based on time series combination model[J]. Electronic Design Engineering, 2023, 31(23): 1–6.

[30] WU CH, FU XL, LI HH, *et al.* Study on inversion of suspended matter in Wuliangsu Lake based on M-GA-BP[J]. Water Resources and Power, 2023(12): 49–52.

[31] DAI YY, GONG SQ, ZHANG CJ, *et al.* Remote sensing model for estimating atmospheric $PM_{2.5}$ concentration in the Guangdong – Hong Kong – Macao Greater Bay Area[J]. Environmental Science, 2024, 45(1): 8–22.

[32] LI QL, LIU F, LIAO W, *et al.* Applicability analysis and regional difference test of CLDAS and CMPAS in Chongqing[J]. Plateau and Mountain Meteorology Research, 2023, 43(1): 119–127.

[33] ZHOU HZ, LIAO P, YANG H, *et al.* The applicability of numerical model and machine learning for near-surface ozone simulation in Lanzhou City[J]. China Environmental Science, 2024(1): 15–27.

[34] GUO Y, ZHANG WQ, LIU L, *et al.* Quantifying the uncertainty sources of future climate projections and narrowing uncertainties with bias correction techniques in Tibetan Plateau[J]. Journal of Beijing Normal University (Natural Science), 2024(1): 87–98.

[35] QIAO JR, YUAN XP, LIANG XD, *et al.* Application of agglomerative hierarchical clustering method in precipitation forecast assessment[J]. Journal of Arid Meteorology, 2022, 40(4): 690–699.

[36] KONG XR, CHEN M, CHEN HR, *et al.* Variation characteristics and influence factors of surface ozone concentration in Lanzhou in 2018–2019[J]. Environmental Engineering, 2022, 40(7): 69–75, 152.

[37] FU CB, XU WS, DAN L, *et al.* Impacts of precursors and meteorological factors on ozone pollution in Hainan Province[J]. Environmental Science &Technology, 2020, 43(7): 45–50.

[38] FU CB, XU WS, DAN L, *et al.* Temporal and spatial variation in ozone and its causes over Hainan Province from 2015 to 2020[J]. Environmental Science, 2022, 43(2): 675–685.

[39] SU XT, FENG J, AN H, *et al.* Trends analysis of fine particulate matter and ozone pollution in typical cities in the Beijing – Tianjin – Hebei region during 2015–2021[J]. Chinese Journal of Atmospheric Sciences, 2023, 47(5): 1641–1653.

[40] WANG XL, HUANG R, ZHANG WX, *et al.* Forecasting ozone and $PM_{2.5}$ pollution potentials using machine learning algorithms: A case study in Chengdu[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2021, 57(5): 938–950.