https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

Influencing factors analysis and prediction of near-surface ozone in Henan Province from 2015 to 2022

Haoming Bao[b], Jiandong Shang[a], Jinzhu Li[b], Gang Wu[a], Haitao Wei[b], Lingling Wang[c], Nan Wang[d], Jingye Shi[e], Wenge Zhou[b], Feng Chen[b], Jiahui Guo[b], Jinyang Wang[b], Dujuan Zhang[a,*], Hengliang Guo[a,*]

[A] *National Supercomputing Center in Zhengzhou,Zhengzhou University，Zhengzhou,450001,Henan,China*

[B] *School of Earth Science and Technology, Zhengzhou University,Zhengzhou,450001,Henan,China*

[C] *Henan Ecological Environmental Monitoring and Safety Center, Zhengzhou 450001, China*

[D] *College of Global Change and Earth System Science, Faculty of Geographical Science, Beijing Normal University,Beijing 100875, China*

[E] *Beijing Key Laboratory of Security and Privacy in Intelligent Transportation, Beijing Jiaotong University, Beijing 100044, China*

* Corresponding author.E-mail address:duduzdj@zzu.edu.cn

* Corresponding author.E-mail address:guohengliang@zzu.edu.cn

Abstract: This study analyzed factors influencing near-surface ozone (O_3) in Henan Province from 2015 to 2022 using real-time pollutant data from the China National Environmental Monitoring Centre and daily meteorological data from the Henan Provincial Ecological Environment Monitoring and Safety Center. Regression and machine learning models (including multiple linear regression (MLR), support vector machine (SVM), random forest (RF), ridge regression (RR), BP neural network, and extreme gradient boosting (XGBoost)) were used to predict O_3 concentrations. The results showed that among the major pollutants (CO, NO_2, SO_2, $PM_{2.5}$, and PM_{10}), there was a consistent negative correlation with O_3. Notably, NO_2 had the strongest negative correlation (r = -0.825), while PM_{10} showed the weakest (r = -0.687). From the perspective of meteorological factors, temperature showed a strong positive correlation with O_3, while wind speed, relative humidity, and precipitation showed weak negative correlations, influencing regional variations in O_3 concentrations. Among the six prediction models constructed to predict O_3 concentrations, the most accurate model for predicting concentrations for the next day was the extreme gradient boosting (XGBoost) model (R^2 = 0.883). For the next 3 days, the

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

random forest (RF) model demonstrated the highest accuracy ($R^2 = 0.704$). Similarly, the random forest model (RF) also exhibited the highest accuracy for predicting the next 7 days ($R^2 = 0.651$). In summary, over the past 7 years, there has been a strong correlation observed between O_3 concentration and other major pollutants, as well as meteorological factors in Henan Province. Therefore, it is essential to implement targeted measures for O_3 pollution prevention and control based on specific weather conditions.

Keywords: Ozone (O_3); Henan Province; Relevance; Prediction model

1 Introduction

Near-surface ozone (O_3) in cities is a secondary pollutant generated by complex photochemical reactions of precursors such as NOx and volatile organic compounds (VOCs), emitted by human activities under specific atmospheric conditions (Liang et al., 2019; Wang et al., 2012). It is one of the major pollutants in the atmosphere. As a strong oxidant, O_3 can enhance atmospheric oxidation capacity, promote the degradation of primary pollutants, and increase the generation rate of secondary pollutants (Feng et al., 2019), thus severely damaging regional and urban air quality (Wang et al., 2017). High concentrations of ground-level O_3 can harm the health of humans, animals, and plants (Ruan et al., 2019). Increased environmental O_3 concentrations lead to changes in biodiversity, decreased productivity, reduced carbon sequestration capacity, and decreased water resources (Feng et al., 2015). These changes have direct negative impacts on human health (Han et al., 2018; Lu et al., 2020), ecosystems (Huang et al., 2019), and economic development (Jakovljević et al., 2021). Therefore, near-surface O_3 pollution has become a highly concerning topic in atmospheric environmental science.

In recent years, with the acceleration of urbanization, photochemical pollution, particularly with O_3 as the main indicator, has become increasingly severe in urban areas of China (Wu et al., 2017). Near-surface O_3 characterized by its reactivity and a

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

wide range of sources, experiences concentration changes that are easily affected by natural conditions and changes in the concentration levels of other pollutants (Jaén et al., 2021). Current studies suggest that the spatiotemporal variation of O_3 concentration is influenced by a combination of meteorological conditions, precursor emissions, and chemical reactions (Lin et al., 2022). In terms of meteorological factors, studies have documented the effects of temperature (Yang et al., 2021), wind speed (Li et al., 2018), relative humidity (Ma et al., 2021), and precipitation (Su et al., 2015) on O_3 levels. For instance, Wang et al. (2015) analyzed the impact of meteorological conditions on O_3 concentration in Shenyang's ambient air and found that O_3 levels are influenced by multiple factors, including temperature, wind speed, humidity, visibility, and general weather conditions. Similarly, Wang et al. (2011) analyzed the impact of weather conditions on the distribution of near-surface O_3 in Fuzhou, revealing that O_3 concentration is closely related to meteorological factors, significantly positively correlated with temperature, and significantly negatively correlated with relative humidity. Additionally, higher average wind speeds are associated with increased O_3 concentrations. An et al. (2009) analyzed the impact of meteorological factors on O_3 in Beijing and found that O_3 concentration is inversely proportional to air pressure, humidity, and visibility, but directly proportional to wind speed and temperature. Studies on precursors such as NOx (Li et al., 2021; Gu et al., 2020) and volatile organic compounds (VOCs) (Wu et al., 2017, Liu et al., 2021), have also been conducted. For instance, Chang et al. (2021) analyzed the driving forces behind O_3 pollution in Shanghai and discovered that reducing VOC emissions can effectively improve O_3 pollution in the city. In terms of O_3 concentration prediction, numerous studies have used the atmospheric chemical transport model and empirical statistical model to estimate near-surface O_3 concentration (Wang et al., 2013). Recently, machine learning technology has shown promising results in predicting O_3 concentrations. Zhao et al. (2022) estimated near-surface O_3 concentration in mainland China using the XGBoost algorithm. The model was validated using cross-validation and self-modeling validation, achieving R^2 values of 0.871 and 0.955, and RMSE values of 12.8 μg/m^3 and 7.514 μg/m^3, respectively, effectively simulating near-surface O_3 concentration. Prybutok et al. (Victor, 2000) confirmed that artificial neural networks predict O_3 concentrations more accurately than conventional statistical models. The inputs for the developed model included pollutants such as CO, NO_2, and NO, as well as meteorological variables. Studies have shown that artificial neural networks outperform the regression models and the autoregressive integrated moving average (ARIMA) model, achieving the lowest scores in mean absolute deviation (MAD) and root mean square error (RMSE). Sheta et al. (2018) used a different type of artificial neural network called hybrid cycle reservoir with jumps (HCRJ) to predict O_3 concentration in eastern Croatia, specifically in the cities of Osijek and Kopački. The system inputs include meteorological variables, PM_{10} levels, and O_3 concentration from the previous day, showing promising application prospects. Faced with the increasingly severe O_3 pollution, it is necessary to identify the factors contributing to O_3 pollution and make accurate predictions of future O_3 concentrations. This is an important prerequisite for the effective prevention and control of O_3 pollution.

Since 2015, atmospheric O_3 pollution in certain regions of China has shown a consistent upward trend, with regional O_3 pollution intensifying annually (Li et al., 2022; Chen et al., 2022; Wu et al., 2022). The northern region of Henan Province, located in northern China, has particularly experienced severe air pollution in recent years (Ke et al., 2022), attracting considerable domestic and international research attention (Xue et al., 2013). In addition, Henan Province has a dense population and extensive transportation industry (Qi et al., 2020), resulting in significant emissions of pollutants. Despite the problem of O_3 pollution being alleviated through initiatives like the Three-Year Action Plan for Pollution Prevention and Control in Henan Province (2018-2020) issued by the Department of Ecology and Environment of Henan Province (Notice, 2018), the overall pollution situation remains severe. Qi et al. (2020) used the Pearson correlation coefficient to describe the influencing factors of O_3 pollution trends in Henan Province in 2017. They found that hourly O_3 concentrations exhibit negative correlations with CO and NO_2 levels. Moreover, correlations between O_3 MDA8 and meteorological factors (such as sunshine duration, temperature, rainfall, visibility, relative humidity, and wind speed) vary across different seasons and cities. However, there has been a notable absence of modeling and prediction of O_3 pollution in Henan Province. In addition, existing studies (Ge et al., 2022; Yan et al., 2022; Qi et al., 2020; Gao et al., 2021; Guo et al., 2015) have relatively little focus on O_3 pollution in Henan Province, lacking comprehensive multi-year analysis of its influencing factors. There is also a lack of future O_3 concentration predictions for the region. Therefore, conducting research on O_3 pollution in Henan Province holds great significance for the prevention and control of air pollution across the entire Central Plains. This study used real-time monitoring data of environmental air quality in Henan Province from 2015 to 2022 to analyze spatiotemporal changes in O_3 and other pollutants (CO, $PM_{2.5}$, PM_{10}, SO_2, and NO_2). It systematically analyzed correlations between O_3 pollution and meteorological factors and other pollutants over the past eight years. Multiple prediction models were used to predict O_3 concentrations, in order to provide valuable insights for future O_3 pollution governance in Henan Province and improve the effectiveness of pollution control measures.

2 Data and Methods

2.1 Study area

Due to data availability, the study focused on 17 cities in Henan Province (excluding Jiyuan), situated between 31°23' to 36°22'N latitude and 110°21' to 116°39'E longitude. Henan Province, located in central China, exhibits diverse terrain with higher elevations in the west and lower in the east. Some parts of the province are located in the warm temperate zone, while the southern part transitions through the subtropical zone, characterized by a continental monsoon climate. It has complex and diverse characteristics such as four distinct seasons, simultaneous rainfall and heat, and frequent meteorological disasters (Qi et al., 2020). For the convenience of description, Henan Province is divided into five major regions: East Henan, South Henan, West Henan, North Henan, and Central Henan, as detailed in Figure 1.

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

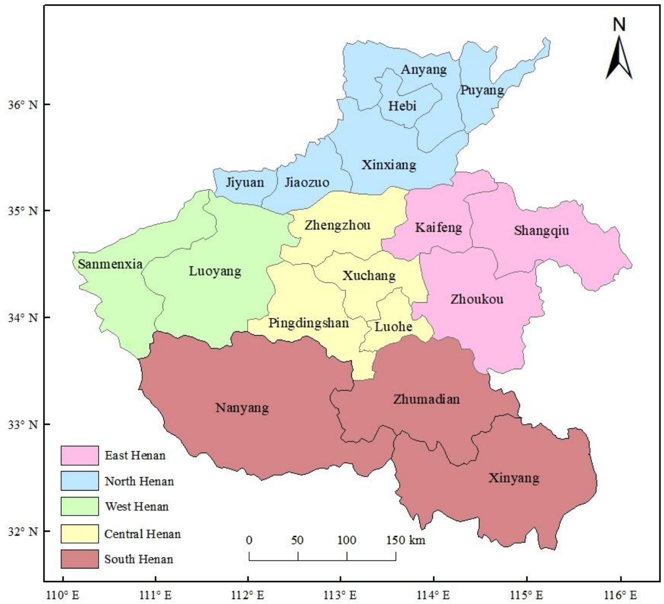

Figure 1. Regional division of Henan Province

2.2 Data source

2.2.1 Pollutant concentration data

Pollutant data (O_3, NO_2, CO, SO_2, $PM_{2.5}$, and PM_{10}) were obtained from the real-time release platform of the China National Environmental Monitoring Centre for national urban air quality (https://quotsoft.net/air/). These data comprise real-time hourly measurements from monitoring stations across 17 prefecture-level cities in Henan Province (excluding Jiyuan). According to the Ambient Air Quality Standards (GB 3095-2012) and the Technical Regulations on Ambient Air Quality Index (on Trial) (HJ633-2012), when analyzing annual changes in air pollutant levels, $\rho(O_3)$ represents the 90th percentile of the daily maximum 8-hour moving average. $\rho(CO)$, $\rho(PM_{2.5})$, and $\rho(PM_{10})$ represent the 95th percentile of the 24-hour average and are denoted as CO-24h-95%, $PM_{2.5}$-24h-95%, and PM_{10}-24h-95%, respectively. $\rho(NO_2)$ and $\rho(SO_2)$ represent the 98th percentile of the 24-hour average and are denoted as NO_2-24h-98% and SO_2-24h-98%, respectively.

2.2.2 Meteorological data

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

The daily meteorological data used in this study, covering the period from January 1, 2015, to December 31, 2021, were sourced from the Henan Provincial Ecological Environment Monitoring and Safety Center. The meteorological data acquired include temperature (TEM, °C), relative humidity (RH, %), average wind speed (WS, m/s), and precipitation (PRE, mm). The obtained data are detailed in Table 1.

Table 1. Experimental data

Data types	Data	Temporal resolution	Unit	Performance form
Pollution Data	O_3	1h	μg / m^3	ρ（O_3-8h-90%）
	CO	1h	μg / m^3	ρ（CO-24h-95%）
	$PM_{2.5}$	1h	μg / m^3	ρ（$PM_{2.5}$-24h-95%）
	PM_{10}	1h	μg / m^3	ρ（PM_{10}-24h-95%）
	NO_2	1h	μg / m^3	ρ（NO_2-24h-98%）
	SO_2	1h	μg / m^3	ρ（SO_2-24h-98%）
Meteorological Data	Temperature	1d	°C	TEM
	Mean Wind Speed	1d	m/s	WS
	Relative Humidity	1d	%	RH
	Precipitation	1d	mm	PRE

2.3 Research methods

2.3.1 Pearson correlation analysis

In data analysis, correlation analysis is a common method used to assess the correlation between two sets of data. The Pearson correlation coefficient, denoted as r and ranging from -1 to 1, is typically used to measure the strength and direction of correlation between two continuous variables. A value of $r \in [-1,1]$, indicates the

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

following: r > 0 indicates a positive correlation, r < 0 indicates a negative correlation, and a larger |r| indicates a stronger correlation, implying a closer correlation between two variables. A value of r = 0 indicates no correlation. The calculation equation is as follows:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{1}$$

where r denotes the correlation coefficient, n represents the number of measurement indicators, X_i indicates the influencing factor, Y_i signifies the O_3 concentration, X stands for the average value of influencing factors, and Y denotes the average O_3 concentration.

2.3.2 Multiple linear regression (MLR) model

The multiple linear regression (MLR) model explores the linear relationship between a dependent variable (dependent variable) and multiple independent variables (explanatory variable). One of its key advantages is its ability to consider the collective impact of independent variables on the dependent variable, thereby revealing comprehensive relationships and improving prediction accuracy. This study organized the obtained data and inputted them into the model. The calculation equation is as follows:

$$\begin{aligned}\rho(\mathrm{O_3\_MDA8}) = {} & a\rho(\mathrm{CO\text{-}24h}) + b\rho(\mathrm{PM_{2.5}\text{-}24h}) + c\rho(\mathrm{PM_{10}\text{-}24h}) \\ & + d\rho(\mathrm{SO_2\text{-}24h}) + e\rho(\mathrm{NO_2\text{-}24h}) + fTEM + gWS \\ & + hRH + iPRE + j + \varepsilon\end{aligned} \tag{2}$$

where a, b, c, d, e, f, g, h, and i denote the regression coefficients for CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2, temperature, wind speed, relative humidity, and precipitation, respectively. $\rho(O_3\_MDA8)$, $\rho(CO\_24h)$, $\rho(PM_{2.5}\_24h)$, $\rho(PM_{10}\_24h)$, $\rho(SO_2\_24h)$, and $\rho(NO_2\_24h)$ represent the daily maximum 8-hour average concentration of O_3 and 24-hour average concentrations of CO, $PM_{2.5}$, PM_{10}, SO_2, and NO_2, respectively. *TEM*, *WS*, *RH*, and *PRE* indicate temperature, wind speed, relative humidity, and precipitation, respectively. j and ε signify the constant term and residual, respectively.

2.3.3 Support vector machine (SVM) model

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

Support vector machine (SVM) is a binary classification model designed to separate samples by finding hyperplane, so that samples from different categories are located on opposite sides of the hyperplane, while maximizing the distance (i.e. interval) to the nearest samples on either side (Cortes et al., 1995). SVM incorporates kernel techniques, enabling it to handle nonlinear classification tasks effectively. Conceptually, SVM solves convex quadratic programming problems by maximizing intervals as a learning strategy. Recognized for its efficacy in classification, SVM excels in solving both linear and nonlinear classification problems, demonstrating good performance across various applications.

2.3.4 Random forest (RF) model

Random forest (RF) is a widely recognized Bagging model based on the decision tree algorithm, which constructs multiple datasets through random sampling and establishes multiple independent decision tree models. The final prediction is obtained by averaging or voting on the predictions from these models (Breiman et al., 2020). RF is versatile, applicable to both classification and regression tasks, known for strong noise resistance and robustness against overfitting [71]. In addition, RF can also perform feature selection and demonstrate high algorithmic efficiency. Therefore, as a machine learning algorithm, RF is widely used in addressing classification and regression problems, with various advantages that can effectively solve many practical problems.

This study used all influencing factors as explanatory variables, with the daily maximum 8-hour average O_3 concentration serving as the dependent variable for constructing an RF regression model. The dataset was divided into training and testing sets, comprising 70% and 30% of the total data respectively, to measure the impact of these factors on O_3 concentration.

2.3.5 Ridge regression (RR) model

Ridge regression (RR) adds a regularization term into the loss function of the least squares method to limit the size of the regression coefficients, thereby preventing overfitting and improving the model's ability to generalize. The regularization parameter, also referred to as the ridge parameter, adjusts the strength

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

of the regularization term in the model. RR is particularly effective in handling high-dimensional and collinear datasets, reducing the impact of multicollinearity and improving the model's stability and prediction ability.

2.3.6 BP neural network model

The BP neural network, proposed by Rumelhart et al. (1986), comprises an input layer, a hidden layer, and an output layer. The hidden layer plays a crucial role in extracting feature information through nonlinear transformations. In different application scenarios, the number of nodes in the hidden layer needs to be carefully chosen and adjusted to match specific data characteristics. In addition, by stacking multiple hidden layers, neural networks can effectively learn complex features, thereby achieving efficient feature representation and classification in diverse scenarios.

2.3.7 Extreme gradient boosting (XGBoost) algorithm

XGBoost is a highly efficient gradient-boosting decision tree algorithm that builds upon GBDT to enhance model performance (Li., 2020). It leverages boosting principles to integrate multiple weak learners into a strong learner, improving model performance through collaborative decision-making across multiple trees and cumulative results accumulation. Widely adopted in practice, XGBoost demonstrated an effective machine learning algorithm.

This study used the gradient boosting algorithm, with nine important parameters set during the model construction process: CO, $PM_{2.5}$, PM_{10}, SO_2, NO_2, temperature, wind speed, relative humidity, and precipitation. Multiple parameters were combined to prevent overfitting during the estimation process and improve the estimation accuracy effectively.

2.4 Experimental setup

For prediction purposes, significant pollutant impact factors and meteorological variables (CO, NO_2, SO_2, $PM_{2.5}$, PM_{10}, temperature, wind speed, relative humidity, and precipitation) were integrated into various statistical regression and machine learning models. Seven years of data from January 1, 2015, to December 31, 2021, were used for training and validation, while data from January 1, 2022, to December

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

31, 2022, were used for testing O_3 concentration predictions in Henan Province over 1, 3, and 7 days. These factors were used as input variables across different models, which were divided into training and testing sets in different proportions. After data cleaning, invalid data were removed, and missing values were filled. Model outputs were compared with real-time O_3 concentration data from monitoring stations to evaluate the performance of the model.

Six regression models were constructed: MLR, SVM, RF, RR, BP neural network, and XGBoost. Their accuracies were evaluated and compared based on their predictive performance.

2.5 Evaluation indicator

In order to test the prediction accuracy of each model, four indicators—root mean square error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), and mean absolute percentage error (MAPE)—were used to evaluate the prediction results on the test set. The calculation equations are as follows:

$$RMSE=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i-\hat{y}_i)^2} \tag{3}$$

$$R^2=1-\frac{(\hat{y}_i-y_i)^2}{\sum_{i=1}^{m}(\bar{y}-y_i)^2} \tag{4}$$

$$MAE=\frac{1}{m}\sum_{i=1}^{m}\left|\hat{y}_i-y_i\right| \tag{5}$$

$$MAPE=\frac{100\%}{m}\sum_{i=1}^{m}\left|\frac{\hat{y}_i-y_i}{y_i}\right| \tag{6}$$

where $\hat{y}$ denotes the predicted value output by each prediction model, y_i represents the true value input for each model, $\bar{Y}$ indicates the average of the true values, and m signifies the number of samples evaluated.

3 Results and discussion

3.1 The interannual variation trend of major pollutants

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

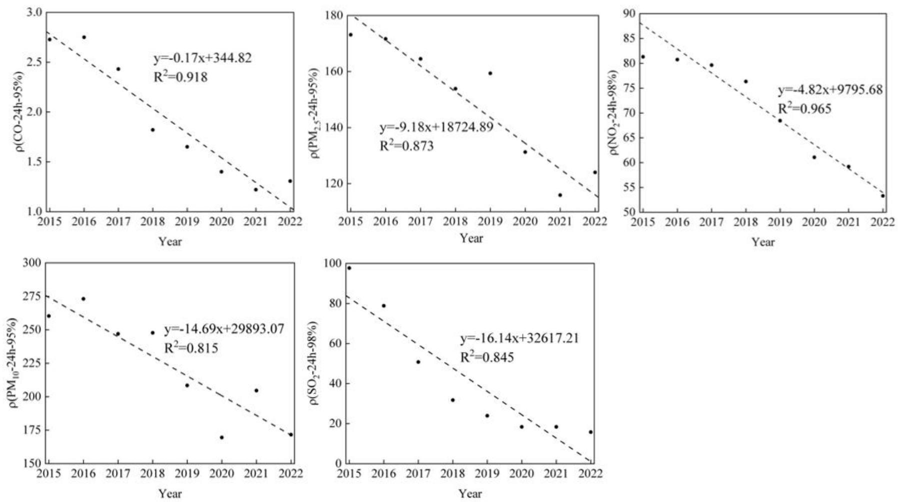

Figure 2. Annual trends of other major atmospheric pollutants in Henan Province from 2015 to 2022

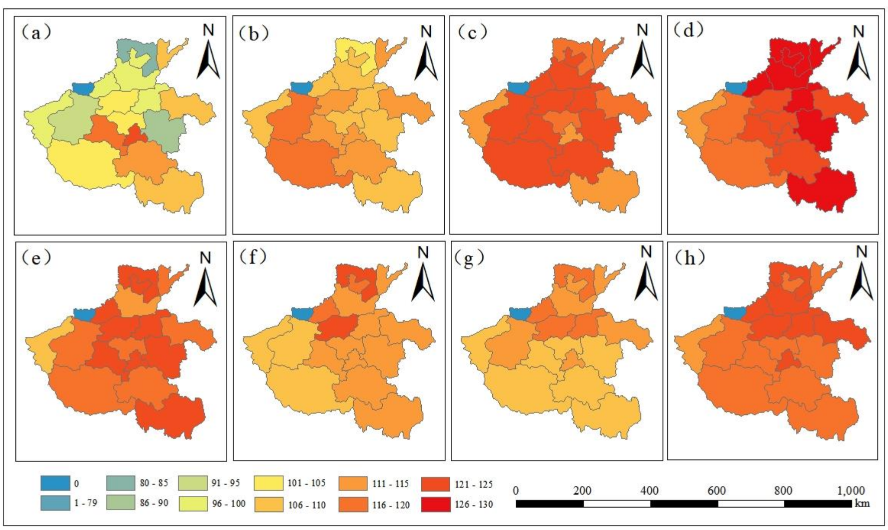

Figure 3. Spatial distribution of annual average O_3 concentration in Henan Province from 2015 to 2022. (a) - (h) denote 2015 to 2022, respectively

The trend of main pollutants in Henan Province is shown in Figure 2. From 2015 to 2022, the 95th percentile of the 24-hour average ρ(CO), ρ($PM_{2.5}$), and the 98th percentile of the 24-hour average ρ(NO_2), ρ(SO_2), and ρ(PM_{10}), showed an overall

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

decrease, with deceleration rates of 0.18 μg/(m^3·a), 6.14 μg/(m^3·a), 3.51 μg/(m^3·a),10.24 μg/(m^3·a), and 11.08 μg/(m^3·a), respectively. The long-term trend of the 90th percentile of ρ(O_3-MDA8) was negatively correlated with the long-term trends of ρ(CO-24h-95%), ρ($PM_{2.5}$-24h-95%), ρ(PM_{10}-24h-95%), ρ(NO_2-24h-98%), and ρ(SO_2-24h-98%). This is because a decrease in ρ($PM_{2.5}$) leads to a decrease in the atmospheric deposition of peroxides (HO_2·) (Li et al., 2020), thereby promoting the generation of O_3. Additionally, a decrease in NO_2 emissions can lead to a decrease in O_3 consumption (Xing et al., 2010). Compared with the concentration data of six pollutants in China from 2013 to 2018 provided by the China Air Quality reanalysis dataset, the concentration of O_3 pollutants in the central region of China showed an overall upward trend (Kong et al., 2021), ranging from 2.3μg/(m^3·a) to 5.4μg/(m^3·a). It indirectly reflects that photochemical pollution in China has increased, while other pollutants show a trend of decreasing year by year. Since this data set can accurately reflect the changing trend of gaseous air pollutants on the surface of China, and this study is consistent with this data set, it also proves that the increase of O_3 concentration is related to the decrease of concentration of other pollutants to a certain extent.Therefore, it is necessary to consider the impact of particulate matter emissions on O_3 concentration and aim to reduce O_3 levels while controlling other major pollutants to improve environmental quality.

Figure 3 illustrates that the high-value area of O_3 concentration is located in central Henan in 2015, namely Pingdingshan City and Luohe City, where the concentrations are 117μg / m^3 and 121μg / m^3 respectively, while the low-value area is located in Anyang City, where the concentrations are 83μg / m^3. In 2018, the O_3 concentration in the province generally increased, and the concentration in the province was between 118.1-129.2μg / m^3. The eastern, southern and northern areas of Henan Province, where the original concentration was low, became high concentration areas, among which the northern area had the highest O_3 concentration. The O_3 concentrations in Anyang City, Jiaozuo City, Hebi City, Xinxiang City and Puyang City were 126μg / m^3, 128μg / m^3, 127μg / m^3, 129μg / m^3 and 128μg / m^3, respectively. In 2019-2021, the concentration of O_3 in all cities decreased slightly, and

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

in 2022, the concentration of O_3 in all cities showed a trend of recovery. On the whole, the distribution area of high O_3 concentration gradually expands, and the O_3 concentration in all cities tends to be the same, rising first and then decreasing, and then rising again. Combined with the HrSOD data set, the overall trend of O_3 concentration in central Henan Province from 2015 to 2020 was selected from the detailed and accurate plot of hourly surface ozone data in China from 2005 to 2020 (Zhang et al., 2022). The results of this study were consistent with this trend. It also shows an overall trend of first rising and then falling, and then rising again, which not only proves the effectiveness of this experiment, but also reflects the seriousness of O_3 pollution in Henan Province. In the next ozone pollution prevention and control, more efforts should be made to improve the quality of atmospheric environment. Therefore, taking prefecture-level cities in Henan Province as the research object, not only the O_3 pollution situation of prefecture-level cities was studied, but also the ozone concentration in recent 8 years was analyzed in time and space to clarify the correlation degree of O_3 concentration among cities and accurately analyze the current problems faced by O_3 pollution in Henan Province.

3.2 The impact of other major pollutants on O_3

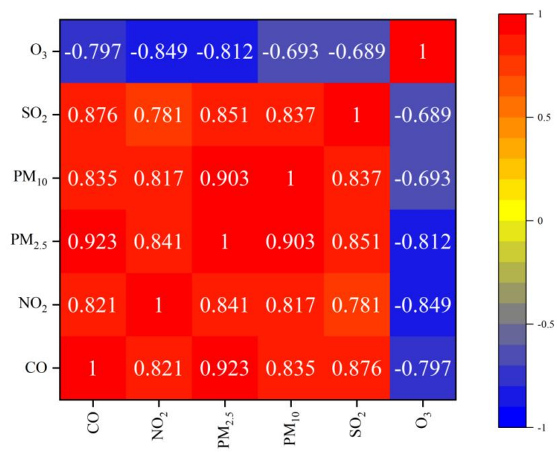

Figure 4. Heat map of correlation coefficients between other major pollutants and O_3 from 2015 to 2022

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

The Pearson correlation coefficient heatmap between O_3 and other major pollutants from 2015 to 2022 is shown in Figure 4. The figure reveals a significant negative correlation between O_3 and other major pollutants, with CO, NO_2, and SO_2 having the largest impact coefficients (-0.791 and -0.825, respectively). Near-surface O_3, which does not have a direct emission source, is categorized as a secondary atmospheric pollutant (Lin et al., 2007). It mainly originates from photochemical reactions and the contribution of precursors such as CO, NO_2, and SO_2 (Li et al., 2019). At high temperatures, atmospheric instability increases, enhancing pollutant diffusion and atmospheric visibility, which in turn increases radiation levels. With the enhancement of solar radiation, photochemical reactions intensify, and O_3 precursors, especially CO, $NO_{2,}$ and SO_2, are more likely to react and generate O_3, under these conditions (Yang et al., 2016). Consequently, CO, NO_2, and SO_2 concentrations decrease while O_3 levels rise. The correlation coefficients between O_3 and $PM_{2.5}$ and PM_{10} were -0.8 and -0.687, respectively. The secondary components in O_3, $PM_{2.5}$, and PM_{10} are all produced through atmospheric chemical reactions, with similar precursors such as NOx and VOCs. O_3 generated by photochemical reactions is affected by the light flux. However, the increase in optical thickness caused by $PM_{2.5}$ and PM_{10} obstructs sunlight, reducing the amount of light reaching the ground and thereby decreasing the rate of photolysis reactions, which leads to a reduction in the generation of O_3.Kong et al. made an important analysis of China's air quality reanalysis data set, and after excluding the disturbance factors of outliers, they concluded that O_3 and other pollutants showed a negative correlation trend in general (Kong et al., 2021).Therefore, as $PM_{2.5}$ and PM_{10} pollution intensifies, sunlight obstruction increases, leading to relatively lower O_3 concentrations (Wang et al., 2019). Therefore, effective atmospheric pollution control measures should focus on reducing precursor substances, which will not only help reduce particulate matter, but also control the formation of near-surface O_3 pollution.

The influence of other pollutants, meteorological conditions and other factors on O_3 pollution in Henan Province was comprehensively explored, and then the ozone

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

concentration was predicted, in order to provide a scientific basis for future ozone concentration pollution in Henan Province, so as to provide a better reference for the prevention and control policy of Henan Province.

3.3 The influence of meteorological factors on O_3

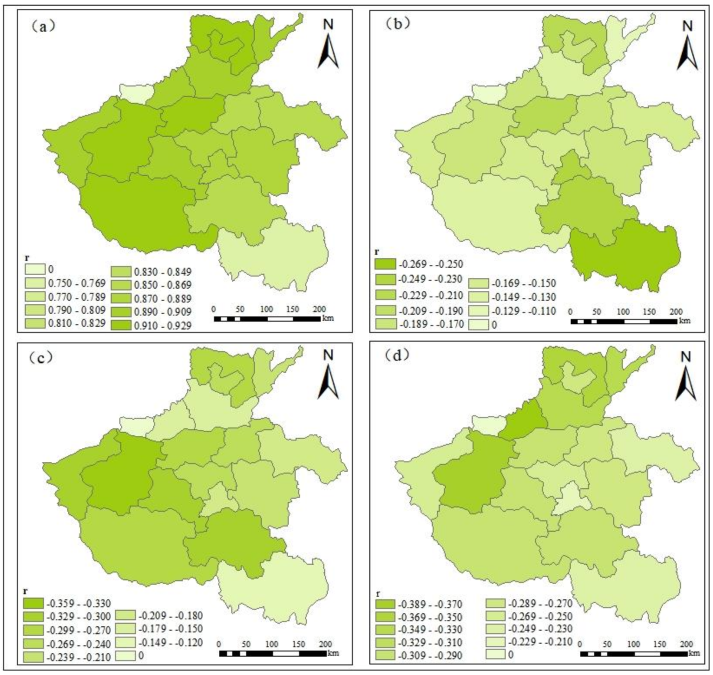

Figure 5. Spatial distribution of correlation coefficients between meteorological factors and O_3 in Henan Province from 2015 to 2022. Meteorological factors represented by (a) - (d) denote temperature, wind speed, relative humidity, and precipitation, respectively.

Table 2. Correlation coefficients between meteorological factors and O_3 in Henan Province from 2015 to 2021

Year	Concentration /（μg / m^3）	Correlation of meteorological factors			
		TEM	WS	RH	PRE
2015	102	0.836**	-0.185*	-0.214*	-0.339**
2016	111	0.832**	-0.162*	-0.233**	-0.326**

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

2017	120	0.845**	-0.198*	-0.226*	-0.295*
2018	124	0.816**	-0.149*	-0.248**	-0.304**
2019	120	0.864**	-0.191*	-0.281**	-0.246*
2020	114	0.833**	-0.201**	-0.320**	-0.305**
2021	112	0.776**	-0.202*	-0.212*	-0.311**
2022	119	0.852**	-0.199*	-0.293**	-0.329**

Note: ** = $P < 0\ 01$, * = $P < 0.05$.

Meteorological factors significantly affect the photochemical environment, precursor diffusion, and atmospheric circulation, all of which play a crucial role in the formation and transformation of O_3. The relationship between O_3 concentration and meteorological factors in Henan Province from 2015 to 2022 is illustrated in Figure 5 and Table 2. The data shows a significant positive correlation between O_3 concentration and temperature, while the correlations between O_3 concentration and wind speed, relative humidity, and precipitation are weak.

The change in temperature reflects the changes in solar radiation intensity. In 2019, the correlation coefficient between temperature and O_3 was at its peak, reaching r = 0.864. Conversely, by 2021, the correlation coefficient decreased to its lowest point of r = 0.776. Temperature is an important factor affecting O_3 concentration, showing a significant increase as temperatures rise. O_3 is generated through photochemical reactions involving primary pollutants under solar radiation. As solar radiation increases and temperatures rise, the rate of these atmospheric photochemical reactions accelerates, leading to higher O_3 concentrations (Yang et al., 2016). In Henan Province from 2015 to 2022, the highest correlation coefficient between temperature and O_3 was observed in Luoyang, at r = 0.926, while the lowest was in Xinyang, at r = 0.762. Cities in West Henan, North Henan, and Central Henan generally exhibited higher correlations compared to those in East Henan. In particular, cities in North Henan consistently showed correlation coefficients above 0.9. North Henan stands out as a significant area of concern for O_3 pollution within Henan Province, requiring focused attention on managing the problem of excessive O_3

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

pollution levels in high-temperature environments.

The influence of wind speed on O_3 concentration is multifaceted: higher wind speed elevates the atmospheric boundary layer, facilitating the mixing of upper-level O_3 downwards. At the same time, increased horizontal diffusion due to higher wind speed dilutes O_3 to a certain extent. These dual effects occur simultaneously. When wind speed is low, the downward mixing effect of O_3 is stronger than the diffusion, leading to a buildup in O_3 concentration. However, as the wind speed increases, the diffusion effect gradually strengthens, balancing out the mixing effect. Therefore, as the wind speed continues to increase, O_3 concentrations continuously decrease (An et al., 2009). In 2021, the correlation coefficient between wind speed and O_3 was at its peak, with r = -0.202. Conversely, in 2018, this correlation was at its lowest, with r = -0.149. Across Henan Province from 2015 to 2022, the highest correlation coefficient between wind speed and O_3 was observed in Xinyang, with r = -0.256, while Puyang had the lowest correlation at r = -0.113. Overall, South Henan exhibited a strong correlation, while other regions showed weaker correlations. Therefore, addressing O_3 pollution during periods of low wind speed in South Henan warrants particular attention.

Relative humidity impacts O_3 concentration by affecting solar ultraviolet (UV) radiation and promoting chemical reactions, one of the main driving forces for O_3 generation. High relative humidity causes water vapor in the air to scatter and absorb UV radiation, reducing UV light reaching the ground, a critical factor for O_3 production. Consequently, reduced UV radiation can lead to a decrease in the rate of O_3 generation. Under high relative humidity conditions, water vapor in the air containing active free radicals like H and OH promotes their generation and activity. These radicals react with O_3, decomposing it into oxygen molecules and accelerating the decomposition process of O_3. A large amount of water vapor in the air consumes O_3 through chemical reactions, thereby reducing its concentration. Conversely, rising ground temperatures elevate water vapor, forming clouds that reduce the intensity of solar radiation, which is unfavorable for photochemical reactions and O_3 generation (Cheng et al., 2016). Therefore, high humidity conditions are generally unfavorable

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

for the accumulation of O_3. The correlation coefficient between relative humidity and O_3 was highest in 2021 at r = -0.320, and lowest in 2021 at r = -0.212. From 2015 to 2022, the correlation coefficient between relative humidity and O_3 was highest in Xinyang at r = -0.34, and lowest in Puyang at r = -0.129. Overall, West Henan and Central Henan exhibit a strong correlation, while other regions show a weaker correlation, indicating a spatial gradient from northwest to southeast Henan. Therefore, West Henan and Central Henan may experience higher O_3 concentrations in a dry environment.

During rainfall, phenomena such as increased cloud cover, heightened wind speed, and decreased temperatures often occur, which are unfavorable for O_3 generation and accumulation. Additionally, pollutants are prone to wet deposition and elimination from the atmosphere, thereby reducing O_3 concentrations (Yang et al., 2018). The correlation coefficient between precipitation and O_3 was highest in 2015, with r = -0.339, and lowest in 2019, at r = -0.246. The correlation coefficient between precipitation and O_3 in various cities in Henan Province from 2015 to 2022 was highest in Jiaozuo at r = -0.385, and lowest in Shangqiu at r = -0.242. Overall, West Henan and North Henan exhibited a strong correlation, while East Henan and South Henan demonstrated a weaker correlation, indicating a spatial gradient from northwest to southeast Henan. Precipitation in Henan Province mostly occurs in summer and autumn, with higher amounts in East Henan and South Henan compared to West Henan and North Henan, especially in Xinyang, South Henan. This phenomenon also explains the occurrence of high relative humidity and lighter O_3 pollution in Xinyang.

From the perspective of meteorological conditions influencing O_3 pollution, temperature, wind speed, relative humidity, and precipitation emerge as pivotal factors. They can serve as preliminarily meteorological indicators for predicting high O_3 concentrations. During high temperatures, low wind speeds, low humidity, and minimal rainfall, the probability of photochemical pollution events characterized by high O_3 concentration is heightened, demanding sufficient attention. Under such weather conditions, specific measures to prevent O_3 pollution should be implemented.

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

3.4 Evaluation of model prediction performance

3.4.1 Multiple linear regression (MLR) prediction performance

The prediction results of the MLR model are shown in Figure 6. As the forecast period increased, the MAE for 1 day, 3 days, and 7 days increased from 15.553 μg/m^3 to 24.506 μg/m^3, marking an increase of 57.56%. Similarly, the RMSE increased from 17.093 μg/m^3 to 28.817 μg/m^3, reflecting an increase of 68.59%. MAPE also increased from 16.931% to 27.292%, representing a 10.361% increase. In contrast, R^2 decreased from 0.824 to 0.535, indicating a decrease of 35.07%. These results indicate that while the MLR model demonstrates strong performance in short-term O_3 concentration prediction, its long-term predictive performance is less effective.

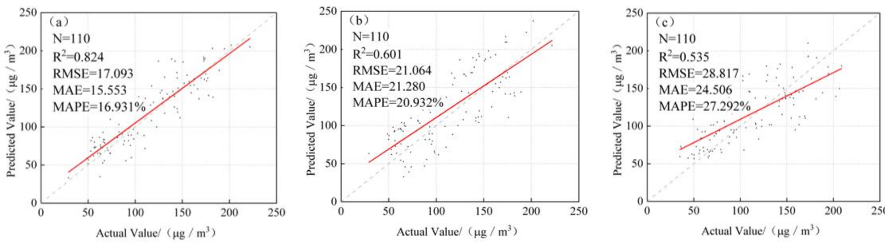

Figure 6. Prediction results of the MLR model. (a) - (c) represent the prediction for the next 1 day, 3 days, and 7 days, respectively.

3.4.2 Support vector machine (SVM) prediction performance

The prediction results of SVM regression are depicted in Figure 7. As the forecast period increased, the MAE for 1 day, 3 days, and 7 days increased from 14.156 μg/m^3 to 21.381 μg/m^3, marking an increase of 51.04%. Similarly, the RMSE increased from 15.894 μg/m^3 to 22.596 μg/m^3, reflecting a 42.17% increase. MAPE also saw an increase from 15.059% to 21.902%, indicating an increase of 6.843%. Meanwhile, R^2 decreased from 0.859 to 0.598, signifying a decrease of 30.38%. These results indicate that while the SVR model performs well in short-term O^3 concentration prediction, its long-term prediction performance is limited.

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

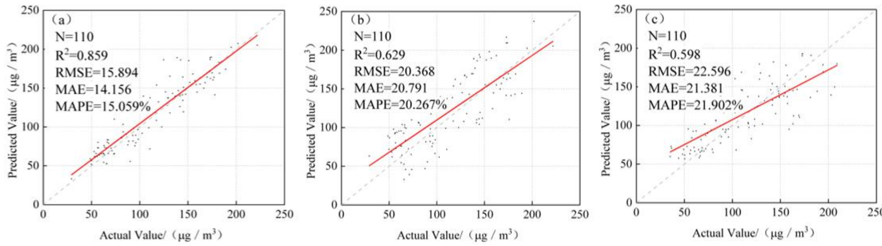

Figure 7. Prediction results of the SVR model. (a) - (c) represent the prediction for the next 1 day, 3 days, and 7 days, respectively.

3.4.3 Random forest (RF) prediction performance

The prediction results of RF regression are shown in Figure 8. As the forecast period increased, the MAE for 1 day, 3 days, and 7 days increased from 12.338 μg/m^3 to 20.378 μg/m^3, representing an increase of 65.16%. Similarly, the RMSE increased from 14.037 μg/m^3 to 23.297 μg/m^3, marking an increase of 65.97%. MAPE also saw an increase from 13.204% to 22.411%, indicating an increase of 9.207%. Meanwhile, R^2 decreased from 0.877 to 0.651, signifying a decrease of 25.77%. These results indicate that the RF model performs well overall in predicting O^3 concentration in both the short and long term.

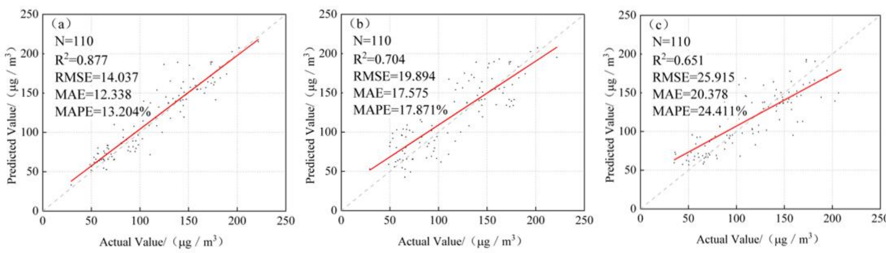

Figure 8. Prediction results of the RF Model. (a) - (c) represent the prediction for the next 1 day, 3 days, and 7 days, respectively.

3.4.4 Ridge regression (RR) prediction performance

The prediction results of RR are shown in Figure 9. As the forecast period increased, the MAE for 1 day, 3 days, and 7 days increased from 16.691 μg/m^3 to 27.001 μg/m^3, representing an increase of 61.77%. Similarly, the RMSE increased

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

from 17.894 μg/m^3 to 31.138 μg/m^3, marking an increase of 74.01%. MAPE also increased from 13.204% to 22.411%, indicating an increase of 11.408%. Meanwhile, R^2 decreased from 0.812 to 0.467, signifying a decrease of 42.49%. These results indicate that the RR model performs poorly overall in predicting O^3 concentration in both the short and long term.

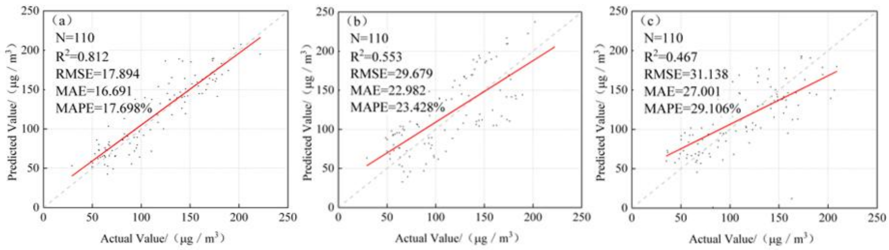

Figure 9. Prediction results of the RR model. (a) - (c) represent the prediction for the next 1 day, 3 days, and 7 days, respectively.

3.4.5 BP neural network prediction performance

The prediction results of the BP neural network model are shown in Figure 10. As the forecast period increased, the MAE for 1 day, 3 days, and 7 days increased from 14.452 μg/m^3 to 24.163 μg/m^3, marking an increase of 67.19%. Similarly, the RMSE increased from 15.894 μg/m^3 to 28.043 μg/m^3, indicating an increase of 76.44%. MAPE also increased from 15.139% to 26.684%, representing an increase of 11.545%. Meanwhile, R^2 decreased from 0.846 to 0.541, signifying a decrease of 36.05%. These results indicate that while the BP neural network model performs well in predicting O_3 concentration in the short term, its long-term prediction performance is poor.

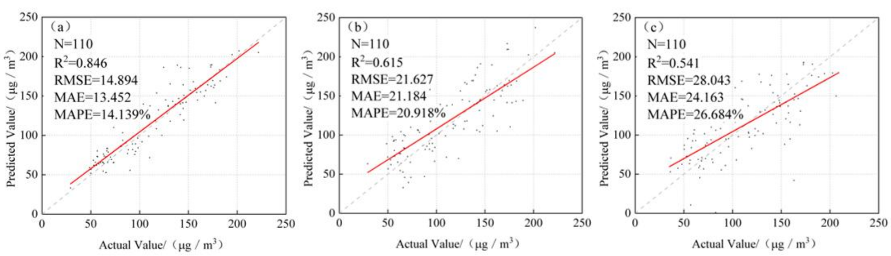

Figure 10. Prediction results of the BP neural network model. (a) - (c) represent predictions for the

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

next 1 day, 3 days, and 7 days, respectively.

3.4.6 Extreme gradient boosting (XGBoost) prediction performance

The prediction results of the XGBoost model are shown in Figure 11. As the forecasting period extended, the MAE for 1 day, 3 days, and 7 days increased from 11.944 μg/m^3 to 20.287 μg/m^3, marking an increase of 69.85%. Similarly, the RMSE increased from 13.579 μg/m^3 to 23.019 μg/m^3, indicating an increase of 69.52%. MAPE also increased from 12.519% to 21.929%, reflecting an increase of 9.41%. Meanwhile, R^2 decreased from 0.883 to 0.621, representing a decrease of 29.67%. These results indicate that the XGBoost model performs well overall in predicting O_3 concentration in both the short and long term.

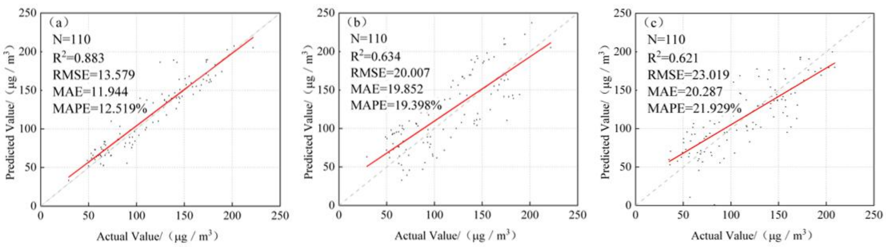

Figure 11. Prediction results of the XGBoost model. (a) - (c) represent predictions for the next 1 day, 3 days, and 7 days, respectively.

3.5 Comparison of accuracy of different prediction models

Table 3a shows a comparison of prediction indicators across various models for the next 1 day, 3 days, and 7 days. Following multiple experiments, optimal parameters were selected for each model, resulting in relatively high accuracy in short-term O_3 concentration prediction. However, over time, the R^2 of each model decreased to varying degrees, with RR exhibiting the largest decline, decreasing by 0.345 from 1 day to 7 days, while RF showed the smallest decline, decreasing by 0.226 over the same period. In prediction for the next day, model accuracy ranked from highest to lowest as XGBoost > RF > SVR > BP neural network > MLR > RR. Specifically, the XGBoost model has the highest accuracy, with metrics of 11.944 (MAE), 13.579 (RMSE), 12.519% (MAPE), and 0.883 (R^2). In the predictions for the

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

next 3 days accuracy ranked RF > XGBoost > SVR > BP neural network > MLR > RR. The RF model achieved the highest accuracy, with metrics of 17.575 (MAE), 19.894 (RMSE), 17.871% (MAPE), and 0.704 (R^2). Similarly, in predictions for the next 7 days, accuracy ranked RF > XGBoost > SVR > BP neural network > MLR > RR. Here, the RF model also exhibited the highest accuracy, with metrics of 20.378 (MAE), 23.297 (RMSE), 22.411% (MAPE), and 0.704 (R^2). In conclusion, both the RF and XGBoost models demonstrated superior accuracy, indicating their capability to effectively simulate O_3 concentration and their practical applicability. Conversely, the RR model showed poorer model accuracy in these predictions.In summary, when predicting short-term O_3 concentration, other models show the limitation of high error, and XGBoost model and RF model should be given priority.

Table 3. Indicator comparison of prediction for the next 1, 3, and 7 days by each model

Future days	Model	MAE	RMSE	MAPE(%)	R^2
1d	MLR	15.553	17.093	16.931	0.824
	SVR	14.156	15.894	15.059	0.859
	RF	12.338	14.037	13.204	0.877
	RR	16.691	17.894	17.698	0.812
	BP Neural Network	14.452	15.894	15.139	0.846
	XGBoost	11.944	13.579	12.519	**0.883**
3d	MLR	21.280	21.064	20.932	0.601
	SVR	20.791	20.368	20.267	0.629
	RF	17.575	19.894	17.871	**0.704**
	RR	22.982	29.679	23.428	0.553
	BP Neural Network	21.184	21.796	20.918	0.615
	XGBoost	19.852	20.007	19.398	0.634
7d	MLR	24.506	28.817	27.292	0.535
	SVR	21.381	22.596	21.902	0.598

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

RF	20.378	23.297	22.411	**0.651**
RR	27.001	31.138	29.106	0.467
BP Neural Network	24.163	28.043	26.684	0.541
XGBoost	20.287	23.019	21.929	0.621

3.6 Discussion

Based on previous studies, this study analyzed the temporal and spatial distribution characteristics of ozone pollution and the characteristics of ozone influencing factors in Henan Province from 2015 to 2022, and made short-term predictions of ozone concentration in the future, and obtained some conclusions about ozone pollution in Henan Province, but there are still some limitations, which need to be further improved and discussed in future studies.

(1) From the perspective of pollutant concentration and meteorology, this paper comprehensively analyzed the spatio-temporal distribution pattern of ozone pollution in Henan Province from 2015 to 2022 and the response of influencing factors, and summarized the ozone pollution situation in Henan Province in the past eight years.

(2) Multiple prediction models such as time series analysis and machine learning were used to make short-term prediction of ozone concentration in Henan Province. The effectiveness of different prediction models was comprehensively evaluated and the optimal model was selected to provide scientific basis for the prevention and control of ozone pollution in Henan Province.

(3) Establish a complete ozone pollution research process from the study of space-time characteristics of ozone, to the analysis of ozone-related factors, to the prediction of ozone concentration, so as to provide a reference for the prevention and control of ozone and other atmospheric pollutants in Henan Province.

(4) This study only analyzed the temporal and spatial distribution characteristics of ozone pollution and the characteristics of ozone influencing factors in Henan Province from the perspective of each city station. The influencing factors were other pollutant factors and meteorological factors, etc. However, ozone pollution may also

be affected by other aspects. In the future research, we can analyze the ozone pollution in Henan Province from the aspects of anthropogenic source and biological source according to the station information of each city.

(5) In this study, only some machine learning models were used to predict short-term ozone concentration in the future. In subsequent studies, machine learning models can be trained, new models can be tried, and model parameters can be continuously optimized to achieve accurate prediction of long-term ozone concentration in Henan Province in the future, which is an important direction for future research.

Ozone is related to the secondary transformation of atmospheric pollutants, which is also the theoretical basis for ozone monitoring and management. In order to better reduce ozone pollution, the monitoring technology system should be improved to realize the transformation from quantitative model to high-quality monitoring based on quality and efficiency. In the background of the known serious ozone pollution problem in Henan Province, appropriate methods should be selected to make predictions, and the whole people should participate in protecting the atmospheric environment.

4 Conclusions

In recent years, O_3 pollution in certain regions of China has been increasing annually, exacerbating regional air quality challenges. The problem of O_3 concentration in Henan Province has also become increasingly prominent. In response to national directives, efforts to prevent and control air pollution are continuously promoted, focusing on strengthening emission reduction measures and effectively managing pollutant sources. Conducting research on O_3 pollution in Henan Province is of great significance for addressing atmospheric pollution in the entire Central Plains region. This study used real-time monitoring data of pollutants and daily meteorological data from Henan Province spanning 2015 to 2022. The Pearson correlation analysis method was used to analyze the correlation between various influencing factors and O_3 concentration in the region. The results indicate that other

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

major pollutants exhibit a negative correlation with O_3, while meteorological factors generally show a weak negative correlation, except for temperature, which exhibits a strong positive correlation. Through the construction of six regression models, adjustment model parameters, and comparison of prediction accuracies, the optimal model for predicting O_3 concentrations was identified. The main contributions of this study are outlined as follows:

(1) The concentrations of other major pollutants in Henan Province, including CO, NO_2, SO_2, $PM_{2.5}$, and PM_{10}, showed an overall downward trend from 2015 to 2021, with decrease rates of 0.18 μg/(m^3·a), 6.14 μg/(m^3·a), 3.51 μg/(m^3·a), 10.24 μg/(m^3·a), and 11.08 μg/(m^3·a), respectively. However, from 2015 to 2022, the annual average O_3 concentration in Henan Province showed a fluctuating pattern: an initial increase, followed by a decrease, and then another increase. There was a notable peak in 2018, with subsequent slight declines from 2019 to 2021 and a slight decrease observed in 2022. It is necessary to strive to reduce emissions of other major pollutants while concurrently achieving a reduction in O_3 concentration to improve environmental quality.

(2) O_3, as a secondary atmospheric pollutant, is mainly affected by precursor emissions of NOx, VOCs, and other particulate matter components. From 2015 to 2022, a significant negative correlation was observed between O_3 and other major pollutants in Henan Province. Among these, NO_2 exhibited the strongest correlation, with a coefficient $r = -0.825$, while PM_{10} showed the weakest correlation, with a coefficient $r = -0.687$. Due to the unique characteristics of O_3, prevention and control should focus on reducing precursor substances and particulate matter.

(3) The concentration of O_3, a regional pollutant, is heavily affected by meteorological conditions. Factors such as temperature, wind speed, relative humidity, and precipitation play pivotal roles. In Henan Province from 2015 to 2022, there was a significant positive correlation between O_3 concentration and temperature, while correlations with wind speed, relative humidity, and precipitation were weak. Temperature exhibited stronger correlations in cities located in West Henan, North Henan, and Central Henan compared to those in East Henan. Wind speed correlations

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

were notably stronger in South Henan but weaker in other regions. Relative humidity showed a spatially decreasing correlation pattern from northwest to southeast of Henan, while precipitation showed a similar gradual transition. O_3 pollution prevention and control plans should be regulated based on local and special weather conditions.

(4) Data from January 1, 2022, to December 31, 2022, were analyzed, and six regression models were developed to predict short-term O_3 concentrations in Henan Province for the next 1 day, 3 days, and 7 days. The results indicated that for predictions of the next day, the models are ranked from highest to lowest as: XGBoost > RF > SVR > BP neural network > MLR > RR. In prediction for the next 3 days, the ranking was RF > XGBoost > SVR > BP neural network > MLR > RR. In prediction for the next 7 days, the models ranked similarly, with RF performing the best, followed by XGBoost, SVR, BP, neural network, MLR, and RR.When predicting short-term O_3 concentration, XGBoost model and RF model are given priority.

CRediT authorship contribution statement

HB: Conceptualization, Formal analysis, Writing - Original Draft, Writing - Review & Editing, Visualization.**JS:** Conceptualization, Formal analysis, Writing - Review & Editing, Visualization, Methodology, Investigation, Funding acquisition.**JL:** Conceptualization, Investigation, Formal analysis, Methodology, Writing - Review & Editing.**GW:** Formal analysis, Writing - Review & Editing, Investigation.**HW:** Investigation, Data Curation, Writing - Review & Editing, Visualization.**L W:**Conceptualization, Formal analysis, Writing - Review & Editing.**NW:** Investigation, Methodology, Formal analysis, Writing - Review & Editing.**JS:** Investigation, Methodology, Formal analysis, Writing - Review & Editing.**WZ:** Conceptualization, Formal analysis, Writing - Review & Editing.**FC:** Methodology, Writing - Review & Editing.**JG:** Methodology, Writing - Review & Editing.**JW:** Methodology, Writing - Review & Editing. **DZ:** Conceptualization, Formal analysis, Supervision, Writing - Review & Editing, Visualization.**HG:** Conceptualization, Formal analysis, Supervision, Writing - Review & Editing, Visualization.

Acknowledgments

The authors would like to thank the support from Key Research and Development Special Projects in Henan Province (241111212300);The Major Science and Technology Special Projects in Henan Province (221100210600); The Major Science and Technology Special Projects in Henan Province (201400211000、201400210700、201400210100).

Competing interests

The contact author has declared that none of the authors has any competing interests.

Data availability statement

The dataset can be found in the national urban air quality real-time release platform of the China National Environmental Monitoring Centre(https://quotsoft.net/air/) and Henan Ecological Environmental Monitoring and Safety Center.

References

An, J., Wang, Y., and Sun, Y.: Assessment of ozone variations and meteorological effects in Beijing, Ecology and Environmental Sciences., 18(3):944 - 951, https://doi.org/10.16258/j.cnki.1674-5906(2009)03-0944-08, 2009.

Breiman, L.: Random forests, Machine Learning., 45(1) : 5-32, 2001.

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

Chang, L., He, F., Tie, X., Xu, J., and Gao, W.: Meteorology driving the highest ozone level occurred during mid-spring to early summer in Shanghai, China, Science of the Total Environment., 785:147253, https://doi.org/10.1016/j.scitotenv.2021.147253, 2021.

Chen, Y., Li, H., Karimian, H., Li, M., Fan, Q., and Xu, Z.: Spatio-temporal variation of ozone pollution risk and its influencing factors in China based on Geodetector and Geospatial models, Chemosphere., 302, https://doi.org/10.1016/j.chemosphere.2022.134843, 2022.

Cheng, N., Li, Y., Zhang, D., Chen, T., Wang, X., Huan, N., Chen, C., and Meng, F.: Characteristics of Ozone over Standard and Its Relationships with Meteorological Conditions in Beijing City in 2014, Environmental Science., 2016,37(6):2041-2051, https://doi.org/10.13227/j.hjkx.2016.06.006, 2014.

Cortes, C., and Vapnik, V.: Support-vector networks, Machine Learning., 1995, 20:273~297, https://doi.org/10.1023/A:1022627411411, 1995.

Feng, Z., Hu, E., Wang, X., Jiang, L., and Liu, X.: Ground-level O_3 pollution and its impacts on food crops in China: a review, Environmental Pollution., 2015,199, 42-48, https://doi.org/10.1016/j.envpol.2015.01.016, 2015.

Feng, Z., Marco, A.D., Anav, A., Gualtieri, M., Sicard ,P., Tian, H., Fornasier. F., Tao, F., Guo, A., and Paoletti,E.: Economic losses due to ozone impacts on human health, forest productivity and crop yield across China., Environment International 131,104966, https://doi.org/10.1016/j.envint.2019.104966, 2019.

Gao, D., Yin ,S., Gu, X., Lu, X., Zhang,H., Zhang, R., Wang, L., and Qi,Y.: Emission inventory and characteristics of vehicle air pollutants in Henan Provin

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

ce from 2016 to 2019, Environmental Science., 2021,42(08):3663-3675, https://doi:10.13227/j.hjkx.202011125, 2021.

Ge, Q., Liu, Y., Yang, H., and Guo, H.: Temporal and spatial characteristics and driving factors of $PM_{2.5}$ in Henan Province from 2015 to 2019, Environmental Science., 2022,43(4):1697-1705, https://doi:10.13227/j.hjkx.202108085, 2022.

Gu, Y., Li, K., Xu, J., Liao, H., and Zhou, G.: Observed dependence of surface ozone on increasing temperature in Shanghai, China, Atmospheric Environment., 2020,221:117108, https://doi.org/10.1016/j.atmosenv.2019.117108, 2020.

Guo, Y.: Study on the causes of $PM_{2.5}$ and ozone formation in Luohe City, Urban Geography., 2015(16):213-214, https://doi:10.3969/j.issn.1674-2508.2015.16.177, 2015.

Han, M., Pa, L., and Wang, J.: Correlation analysis of surface ozone with precursors and particles in Urumqi, Environmental Protection Science., 2018,44(4) : 21-26, https://doi:10.16803/j.cnki.issn.1004-6216.2018.04.004, 2018.

Huang, X., Shao, T., Zhao, J., Cao, J., and Song, Y.: Spatio-temporal differentiation of ozone concentration and its driving factors in Yangtze River Delta urban agglomeration, Resources and Environment in the Yangtze Basin., 2019,28(6):1434-1445, https://doi:10.11870/cjlyzyyhj201906018, 2019.

Jaén, C., Udina,M., and Bech, J.: Analysis of two heat wave driven ozone episodes in Barcelona and surrounding region: Meteorological and photochemical modeling, Atmospheric Environment., 2021,246: 118037, https://doi.org/10.1016/j.atmosenv.2020.118037, 2021.

Jakovljević, T., Lovreškov, L., Jelić, G., Anav, A., Popa, I., Fornasier, M.F., Proietti, C., Limić,I., Butorac, L., Vitale, M., and Marco, A.D.: Impact of ground-level ozone on Mediterranean forest ecosystems health, Science of the Total Environment., 2021,783:147063, https://doi.org/10.1016/j.scitotenv.2021.147063, 2021.

Ke, B., He, C., Yang, L., Ye, Z., Yi, J., Tian, Y., Mu, H., Tu, P., Han,C., and Hong, S.: Temporal and spatial distribution characteristics and driving factors of surface ozone in North China, China Environmental Science., 2022,42(04):1562-1574, https://doi:10.3969/j.issn.1000-6923.2022.04.009, 2022.

Kong, L., Tang, X., Zhu, J., Wang, Z., Li, J., Wu, H., and Carmichael, G.R.: A 6-year-long (2013–2018) high-resolution air quality reanalysis dataset in china based on the assimilation of surface observations from CNEMC, Earth System Science Data., 13(2), 529-570, https://doi.org/10.5194/essd-13-529-2021, 2021.

Li, H., Peng, L., Bi, F., Li, L., Bao, J., Li, J., Zhang, H., and Chai, F.: Strategy of Coordinated Control of $PM_{2.5}$ and Ozone in China, Research of Environmental Sciences., 2019, 32(10): 1763-1778, https://doi:10.13198/j.issn.1001-6929.2019.09.18., 2019.

Li, K., Jacob, D.J., Liao, Hong., and Bates, K.H.: Anthropogenic drivers of 2013-2017 trends in summer surface ozone in China, Proceedings of the National Academy of Sciences of the United States of America., 2019,116(2):422-427, https://doi.org/10.1073/pnas.1812168116, 2019.

Li, M., Wang, T., Xie, M., Li, S., Zhuang, B., Fu, Q., Zhao, M., Wu ,H., Liu, J., Saikawa, E., and Liao, K.: Drivers for the poor air quality conditions in North China Plain during the COVID-19 outbreak, Atmospheric Environment., 2021,246:118103, https://doi.org/10.1016/j.atmosenv.2020.118103, 2021.

Li, Q., Wang, J., Liu, X., Dong, J., and Wei, L.: Temporal and Spatial Distrib

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

ution of Ozone and Effects of Meteorological and Environmental Factors on Ozone in the Urban Areas of Lanzhou City. Environmental Protection Science., 2018,44(02):78-84,97, https://doi:10.16803/j.cnki.issn.1004-6216.2018.02.016, 2018.

Li, T., Chen, J., Weng, J., Shen, J., and Gong, Y.: Ozone pollution synoptic patterns and their variation characteristics in Guangdong Province, China Environmental Science., 2022,42(05):2015-2024, https://doi:10.3969/j.issn.1000-6923.2022.05.004, 2022.

Li, Y., Qin, K., Li, D., Fan, W., and He, Q.: Estimation of ground-level ozone concentration based on GBRT, China Environmental Sciences., 2020,40(3):997~1007, https://doi:10.3969/j.issn.1000-6923.2020.03.008, 2020.

Liang, S., Li, X., Teng, Y., Fu, H., Chen, L., Mao, J., Zhang, H., Gao, S., Sun, Y., Ma, Z., and Azzi, M.: Estimation of health and economic benefits based on ozone exposure level with high spatial-temporal resolution by fusing satellite and station observations, Environmental Pollution., 255:113267, https://doi.org/10.1016/j.envpol.2019.113267, 2019.

Lin, C., Wang, Z., Chou, C.C.K., Chang, C.C., and Liu.S.C.: A numerical study of an autumn high ozone episode over southwestern Taiwan, Atmospheric Environment., 2007,41(17) : 3684-3701, https://doi.org/10.1016/j.atmosenv.2006.12.050, 2007.

Lin, W., and Guo, X.: Spatial and temporal distribution characteristics of ozone in Urban agglomerations in China, China Environmental Science., 2022,Vol. 42 ,2022,42(06):2481-2494, https://doi:10.3969/j.issn.1000-6923.2022.06.001, 2022.

Liu, Y., Shen, H., Mu, J., Li, H., Chen, T., Yang, J., Jiang, Y., Zhu,Y., Meng, H., Dong, C., Wang, W., Xue, L.: Formation of peroxyacetyl nitrate (PAN) and

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

its impact on ozone production in the coastal atmosphere of Qingdao, North China, Science of the Total Environment., 2021, 778:146265, https://doi.org/10.1016/j.scitotenv.2021.146265, 2021.

Lu, X., Zhang, L., Wang, X., Gao, M., Li, K., Zhang, Y., Yue, X., Zhang, Y.: Rapid increases in warm-season surface ozone and resulting health impact in China since 2013, Environmental Science & Technology Letters., 2020,7(4):240-247, https://doi.org/10.1021/acs.estlett.0c00171, 2020.

Ma, M., Yao, G., Guo, J., and Bai, K.: Distinct spatiotemporal variation patterns of surface ozone in China due to diverse influential factors, Journal of Environmental Management., 2021,288:112368, https://doi.org/10.1016/j.jenvman.2021.112368, 2021.

Notice of the People's Government of Henan Province on Issuing the Three-year Action Plan for the Battle against Pollution in Henan Province (2018-2020). Bulletin of Henan Provincial People's Government, 2018(19):4-29.

Qi, Y., Yu, S., Yang, J., Yin, S., Cheng, J., and Zhang, R.: Analysis of Characteristics and Meteorological Influence Factors of Ozone Pollution in Henan Province, Environmental Science., 2020,41(2), https://doi:10.13227/j.hjkx.201908122, 2020.

Ruan, Z., Qian, Z., Guo, Y., Zhou, J., Yang, Y., Acharya, B.K., Guo, S., Zheng, Y., Cum-mings-Vaughn, L.A., Rigdon, S.E., Vaughn, M.G., Chen, X., Wu, F., and Lin, H.: Ambient fine particulate matter and ozone higher than certain thresholds associated with myopia in the elderly aged 50 years and above, Environmental Research., 2019,177, 108581,https://doi.org/10.1016/j.envres.2019.108581, 2019.

Rumelhart, D., Hinton, G., and Williams, R.: Leaning internal representations by back-propagating errors, Nature., 1986,323, https://doi:10.1038/323533a0, 1986.

Sheta, A., Faris, H., Rodan, A., Kovač-Andrić, E., and Ala'M. Al-Zoubi.: Cycle reservoir with regular jumps for forecasting ozone concentrations: Two real cases from the east of Croatia: CRJ for forecasting ozone concentrations, Air Quality, Atmosphere & Health., 2018,11:559-569, https://doi: 10.1007/s11869-018-0561-9, 2018.

Su, T., Li, Z., and Kahn R.: Relationships between the planetary boundary layer height and surface pollutants derived from lidar observations over China: regional pattern and influencing factors, Atmospheric Chemistry and Physics., 2018,18(21):15921-15935, https://doi.org/10.5194/acp-18-15921-2018, 2018.

Victor, R., Junsub, Y., and David, M.: Comparison of neural network models with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations, European Journal of Operational Research., 2000,122(1), https://doi.org/10.1016/S0377-2217(99)00069-7, 2000.

Wang, C., Wang, S., Yang, B., Zhang,L., Wang, L., and Liu, M.: Study of the Effect of Meteorological Conditions on the Ambient Air Ozone Concentrations in Shenyang, Environmental Monitoring in China., 2015,31(3):32 - 37, https:// doi:10.3969/j.issn.1002-6002.2015.03.007, 2015.

Wang, H., Lin, C., Chen, X.,Yu, Y., and Bai, L.: Effects of weather condition on surface ozone distribution in FuZhou, Ecology and Environment., 2011,20(8-9):1320 - 1325, https://doi:10.3969/j.issn.1674-5906.2011.08.022, 2011.

Wang, M., Zheng, Y., Liu, Y., Li,Q., and Ding, Y.: Characteristics of ozone a nd its relationship with meteorological factors in Beijing-Tianjin-Hebei Region, China Environmental Sciences., 2019, 39(7): 2689-2698, https://doi:10.3969/j.issn. 1000-6923.2019.07.001, 2019.

Wang, T., Xue, L., Brimblecombe, P., Lam, Y.F., Li, L., Zhang, L.: Ozone pollutionin China: a review of concentrations, meteorological influences, chemical pre-cursors, and effects, Science of the Total Environment, 2017,5751582-1596, https://doi.org/10.1016/j.scitotenv.2016.10.081, 2017.

Wang, Y., Hopke, P.K., Xia, X., Rattigan, O.V., Chalupa, D.C., Utell, M.J.: Source apportionment of airborne particulate matter using inorganic and organic species as tracers, Atmospheric Environment, 2012,55, https://doi.org/10.1016/j.atmosenv.2012.03.073, 2012.

Wang, Z., Li, X., Wang, Z., Wu, X., Che, F., Nie, P.: Application Status of Models-3/CMAQ in Environmental Management, Environmental Science & Technology, 2013,36(S1):386-39, https://doi:10.3969/j.issn.1003-6504.2013.6L.091, 2013.

Wu, K., Wang, Y., Qiao, Y., Liu, Y., Wang, S., Yang, X., Wang, H., Lu, Y., Zhang, X., Lei, Y.: Drivers of 2013–2020 ozone trends in the Sichuan Basin, China: Impacts of meteorology and precursor emission changes, Environmental Pollution, 2022,300, https://doi.org/10.1016/j.envpol.2022.118914, 2022.

Wu, R., Xie, S.: Spatial distribution of ozone formation in China derived from emissions of speciated volatile organic compounds, Environmental Science & Technology, 2017,51(5):2574-2583, https://doi.org/10.1021/acs.est.6b03634, 2017.

Xing, J., Wang, S., Jang, C., Zhu, Y., Hao, J.: Nonlinear response of ozone to precursor emission changes in China: a modeling study using response surface methodology, Atmospheric Chemistry and Physics, 2010,11(10):5027-5044, https://doi.org/10.5194/acp-11-5027-2011, 2010.

Xue, M., Ma, J., Li, Y., Zhu, S., Zhao, B., Yan, P., Ding, G., Jin, R.: Chemical characteristics of air masses from different urban and industrial centers in the Huabei region of China, Atmospheric Environment, 2013,71, https://doi.org/10.1016/j.atmosenv.2013.01.045, 2013.

Yan, Y., Yin, S., He, Q., Qin, K., Zhang, R.: Trend Changes in Ozone Pollution and Sensitivity Analysis of Ozone in Henan Province, Environmental Science, 2022,43(6), https://DOI:10.13227/j.hjkx.202108287, 2022.

Yang, L., Guo, Y., Zhang, J., Wang, W., Bai,J.: Analysis of the Pollution Characteristics and Influencing Factors of the Ozone in Yinchuan City, Environmental Protection Science, 2016,42(2):55-59, https://doi: 10.16803/j.cnki.issn.1004-6216.2016.02.012, 2016.

Yang, L., Xie, D., Yuan, Z., Huang, Z., Wu, H., Han, J., Liu, L., Jia, W.: Quantification of regional ozone pollution characteristics and its temporal evolution: Insights from identification of the impacts of meteorological conditions and emissions, Atmosphere, 2021,12(2):279, https://doi.org/10.3390/atmos12020279, 2021.

Yang, X., Li, X.: A numerical study of photo chemical reaction mechanism of ozone variation in surface layer, Chinese Journal of Atmospheric Sciences, 1999,23(4)：427-438, https://DOI: 10.3878/j.issn.1006-9895.1999.04.06, 1999.

Yang, X., Xue, L., Wang, T., Wang, X., Gao, J., Lee, S., Blake, D.R., Chai, F., Wang, W.: Observations and explicit modeling of summertime carbonyl formation in Beijing: identification of key precursor species and their impact on atmospheric oxidation chemistry, Journal of Geophysical Research-atmospheres, 2018,123,1426-1440, https://doi.org/10.1002/2017JD027403, 2018.

Zhang, W., Liu, D., Tian, H., Pan, N., Yang, R., Tang, W., Yang, J., Lu, F., Dayananda,

https://doi.org/10.5194/egusphere-2024-3495
Preprint. Discussion started: 5 February 2025

B., Mei, H., Wang, S., Shi, H.: Recurrent mapping of Hourly Surface Ozone Data (HrSOD) across China during 2005–2020 for ecosystem and human health risk assessment, Earth System Science Data Discussions, 2022: 1-36, https://DOI:10.5194/essd-2022-428, 2022.