AIMS Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 8(8): 18717–18733.
DOI: 10.3934/math.2023953
Received: 04 December 2022
Revised: 06 May 2023
Accepted: 12 May 2023
Published: 02 June 2023

Research article

Prediction of the air quality index of Hefei based on an improved ARIMA model

Jia-Bao Liu* **and Xi-Yu Yuan***

School of Mathematics and Physics, Anhui Jianzhu University, Hefei 23060, China

* **Correspondence:** Email: liujiabaoad@163.com, yxy17730011211@163.com.

Abstract: With the rapid development of the economy, the air quality is facing increasingly severe pollution challenges. The air quality is related to public health and the sustainable development of the environment of China. In this paper, we first investigate the changes in the monthly air quality index data of Hefei from 2014 to 2020. Second, we analyze whether the Spring Festival factors lead to the deterioration of the air quality index according to the time sequence. Third, we construct an improved model to predict the air quality index of Hefei. There are three primary discoveries: (1) The air quality index of Hefei has obvious periodicity and a trend of descent. (2) The influencing factors of Spring Festival have no significant effect on the air quality index series. (3) The air quality index of Hefei will maintain a fluctuating and descending trend for a period of time. Finally, some recommendations for the air quality management policy in Hefei are presented based on the obtained results.

Keywords: air quality index; univariate regression; prediction model; X-12-ARIMA

1. Introduction

With the development of the economy and technology, the world climate and environment are facing more and more challenges. Thus, countries of the world have to pay more attention to the air pollution. With the gradual globalization of the economy, many countries have begun to advocate for ecological globalization. Air quality has become a concern of many environmentalists at home and abroad. Studies have confirmed that the long-term inhalation of air pollutants increases health risks, such as cardiovascular, respiratory and lung effects [1–5]. In order to improve the technical level of environmental monitoring, modern environmental monitoring technology and equipment are used to monitor the overall air quality and pollutant emissions. It ensures the implementation of air pollution prevention and control to a certain extent. With the use of computer software to examine data, one can control the air quality index (AQI) more effectively. Meanwhile, we can evaluate the effectiveness of existing air control policies, use projected data to correct the existing policies and improve it.

The AQI describes the cleanliness and pollution degree of the air. The U. S. EPA uses five main pollution standards to calculate the air quality: ground ozone, particulate matter pollution, carbon monoxide, sulfur dioxide and nitrogen dioxide. The main factors that affect air quality are vegetation coverage and pollutant emissions. Population urbanization rate, annual average temperature, power consumption and industrial waste gas treatment facilities are strong driving factors, which play a fundamental role in reducing the concentration of pollutants [6]. The influence factors vary in different regions since the different conditions in different cities, such as the promotional effect of the digital economy on urban resilience levels vary significantly across regions [7]. At the beginning of the large-scale spread of the epidemic in 2020, scholars modeled the changes of air quality in various provinces and cities. The results showed that the emissions of primary and secondary pollutants were reduced under the constraints of residents' work and life. The air quality during that period was significantly improved. It means that the residual pollutants have great ramifications on air quality [8, 9].

Since the 21st century, the economy in China has been rapidly developing. The prompt rise of the secondary industry and the acceleration of urbanization made China overtake Germany, Japan and other developed countries hastily, becoming the second largest country in terms of economic output. China has completed more than the 200 years of urbanization and industrialization of developed countries during the past 40 years, which will inevitably bring corresponding air pollution problems. More and more serious haze has appeared in the Yangtze River Delta, Beijing-Tianjin-Hebei and other economic zones.

Scholars most focus on the urban agglomerations, the economic zones and the regional characteristics of the AQI while investigating the air quality and energy consumption of China [10–12]. As one of the constituent provinces of the Yangtze River Delta Economic Zone, Anhui Province has developed rapidly in recent years. As the provincial capital city, Hefei has been included in the scope of the new first-tier cities in 2020. In the past decade, the city's regulated industrial added value has maintained an average annual medium-to-high growth rate of 12.2%, which is nearly 6 percentage points higher than that of the whole country. Industrial investment grew at an average annual rate of 12.1%, which is 6.2 points percentage higher than that of the whole country. While developing the economy, Hefei City is also controlling pollutant emissions, insisting on green and low-carbon technology, leading the development. The overall environmental efficiency development of Hefei was at a higher level than other prefecture-level cities of Anhui Province during 2015–2020 [13]. The energy consumption of industrial added value in the city has decreased by 66.76% in the past decade [14].

It is of great significance to predict the trend of AQI of Hefei. To predict the AQI, scholars have used a BP neural network, grey prediction model and LSTM-network to forecast the index [15], and all of them have achieved ideal prediction results. The autoregressive (AR) model is a model that uses its own historical data as regression variables. The advantage of autoregressive model is that it requires a small amount of data and it is suitable for situations that affected by its own historical factors greatly [16]. The ARIMA model has achieved ideal fitting results when applied to predict air quality-related indicators, such as $PM_{2.5}$, PM_{10}, NO_2, etc. [17–19]. At the same time, the combination of ARIMA model and other models also shows good results in the prediction of air pollutant-related indicators, such as a hybrid model using MODWT and ARIMA, a wavelet-ARMA/ARIMA model and so on [20, 21]. These indicators are significant factors affecting the AQI and important causes of aggravation of air pollution [22–24].

The above cases have some defects when predicted by the ARIMA model, that is, they cannot take into account the impact of special external factors on the data. Some series show obvious periodicity and seasonality, but the period lengths of different series are not the same. If the observation time is not long enough, some periodicity may be missed. Due to the influence of some social and economic development factors, there may be some fixed changes in the time series on some special dates. Therefore, these specific time series usually contain various elements that must be adjusted in order to forecast the data correctly. One of these elements is called the trading day effect (also called the day-of-week effect). Thus, the combination of periodicity and special trading days is likely to have a considerable impact on the time series, making it difficult to analyze the data properly unless these effects are adequately considered [25]. Therefore, in order to grasp the changing trend of the time series more accurately, the periodicity factor in the time series decomposition factor can be redefined as a special trading day factor to improved adaptation to the different characteristics of the series. Regression analysis can be used to determine whether the influence of special trading day factors on time series is significant.The X-11 model uses three different moving average methods to calculate the factorization of time series, and it fits the seasonal adjustment program of time series through the factorization of three stages. To avoid the data loss caused by the moving average, the ARIMA process is used to model the data to supplement the serial values before X-11 processing. On this basis, the same pretreatment of time series is strengthened, which is called the X-12-ARIMA model.

In this paper, we use the new model to analyze the trading days of the AQI of Hefei. The main contents of this paper are as follows: In Section 2.1, we introduce the deterministic factor decomposition of time series. In Section 2.2, we explain the seasonal adjustment model and three moving average models. In Section 2.3, the improved model that considers a special trading day is introduced, and we give the modeling process of X-12-ARIMA completely. In the third section, the X-12-ARIMA model is used to examine whether the influential factors of the Spring Festival will affect the AQI of Hefei or not. The model is used to predict the AQI as well.

2. Methodology

2.1. Deterministic factor decomposition

For deterministic time series, factor decomposition methods are commonly used for analysis. Statisticians believe that all time series can be decomposed into four components: long-term trends, cyclical fluctuations, seasonal variations and random fluctuations. When performing deterministic time-series analysis, the series may contain one of these four influencing factors, or it may be a composite series with a mixture of several components. But, the four factors above can be used to describe all of the time series, meaning that all time series can be fitted with a function as $X = f(T_t, C_t, S_t, I_t)$ [26].

The commonly used functions are additive and multiplicative functions, and the corresponding factor decomposition models are constructed as additive models and multiplicative models. The multiplicative seasonal model of time-series is called a multiplicative seasonal autoregressive differential moving average model. It is a time series model that constructed by introducing the idea of multiplicative seasons based on the basic autoregressive differential moving average model

(ARMA model) [27]. The ARMA model represents the time series model as three parts: difference, autoregression and moving average [28]. The model is often used in series with complex interactions, such as seasonal effects, long-term trends and random fluctuations. Compared with the ARIMA model, it pays more attention to the periodic fluctuation state reflected by the data and the seasonality in the series [29]. The multiplicative model can be expressed as

$$X = T_t \times C_t \times S_t \times I_t.$$

In social and economic life, it is difficult to distinguish cyclical factors and trend factors when the observation period is not long enough. Some socioeconomic phenomena will be significantly affected by some special dates. Based on the multiplicative model, economists improved the deterministic factor decomposition model, changing the cyclical factor to a special trading day factor. The new factors are as follows: long-term trend, seasonal factor, trading day factor and random fluctuation. That means that the time series can be fitted as $X_t = f(T_t, D_t, S_t, I_t)$.

2.2. Seasonal adjustment model

In 1954, Shiskin applied the moving average method to seasonal adjustment, which is called X-1 [30]. After that, Shisskin continuously improved the method and successively developed the seasonal adjustment program from X-3 to X-10. The famous X-11 seasonal adjustment program was launched in 1965, and it has been widely used in the official and commercial departments of the USA because of its excellent adaptability and effectiveness [31].

For the models with obvious seasonal factors, the seasonal factors will cover up the long-term development trend, so it is necessary to decompose the factors when studying the development of socioeconomic phenomena and excluding the influence of seasonal fluctuations. The moving average is often used to eliminate the seasonality of time-series data, and the moving average ratio can effectively extract the seasonal effect. However, the fitting of high-order polynomial functions by a simple moving average is not accurate enough. The X-11 model uses three different moving average methods to calculate the factorization of time series, and it fits the seasonal adjustment program through the factorization of three stages [32].

The X-11 seasonal adjustment model is the most commonly used standard method for statistical and commercial organizations to use for decomposition. The X-11 method was developed by the United States Census Bureau and dates back to the 1950s [33]. The estimated value of the trend period obtained by this method can be used for all observations, including the end point, and it allows the seasonal components to change slowly with time. The X-11 model also has some complex ways to deal with trading day changes, holiday effects and the effects of known predictors. It deals with additive and multiplicative decomposition at the same time, and it robust against outliers and horizontal offsets in time series.

The following are three moving average methods for seasonality adjustment using the X-11 model.

(1) Moving average method

The core of the X-11 program is the moving average method. One of its important features is that it can select functions according to the characteristics of the sequence, such as the number of moving average terms, outliers and so on which can be determined by the program itself [34].

The moving average method is one of the most commonly used smoothing methods. The moving average method can be used to eliminate random fluctuations and seasonal effects, yielding the changing trend of time series. The moving average method is calculated with the Eq (2.1):

$$M\left(x_t = \sum_{i=-k}^{f} \theta_i x_{t-i}\right), \forall k, f > 0, \tag{2.1}$$

where $M(x_t)$ is called the $k + f + 1$ period moving average function of the series x_t and θ_i is called the moving average coefficient or the moving average operator.

(2) Henderson weighted moving average

The simple central moving average can well extract the information of the primary function and the quadratic function when extracting the trend information. But, for the curves with more than a quadratic degree, it is not enough to extract the trend information. The X-11 process needs to further use the Henderson weighted moving average on the basis of the simple moving average [35].

The Henderson weighted moving average means that $S^2 = \sum_{i=-k}^{f} \left(\nabla^3 \theta_i\right)^2$ is minimized under the constraints of $\sum_{i=-k}^{k} \theta_i x_{t-i} = 1$ and $\sum_{i=-k}^{k} i\theta_i = 0$. θ_i is the weighting coefficient of the moving average. Among them, S^2 is equal to the square sum of the third-order difference of the moving average coefficient. It is equivalent to taking a cubic polynomial as an index of smoothness, which requires S^2 to be minimized to make the smoothing value as close to a cubic curve as possible.

(3) Musgrave asymmetric moving average

The above two moving average methods can well eliminate the trend and extract linear or nonlinear trend information, but they are all central moving averages. If the moving average period is $2k+1$, then the moving average fitting will lead to the loss of the front k-period and the last k-period information of the sequence. Therefore, in 1964, the statistician Musgrave constructed the Musgrave asymmetric moving average method to solve this problem to supplement the smooth fitting of the final k-period data [30]. Taking the ratio-to-moving average method as the theoretical basis, a simple treatment of the end value, the asymmetric moving average, is adopted in the X-11 model [31].

The construction idea of the Musgrave asymmetric moving average is that a set of central moving average coefficients is known, which satisfies the premise constraints, such as the minimum variance and optimal smoothness of $\sum_{I=-k}^{k} \theta_i = 1$. Now we need to find another set of non-central moving average coefficients $min\left\{E\left(\sum_{i=-k}^{k} \theta_i x_{t-i} - \sum_{i=-(k-d)}^{k} \varphi_i x_{t-i}\right)\right\}^2$, where d is the number of terms for supplementary smoothing. This coefficient set also satisfies the constraint $\sum_{i=-k}^{k-d} \varphi_i = 1$ with a sum of 1, and its fitting value can be infinitely close to the fitting value of the central moving average. That is, the modification to the existing estimated value of the central moving average is minimal.

With this guiding idea, Musgrave applied the concept of noise-to-signal ratio $R = \frac{\bar{I}}{\bar{C}}$ to calculate the coefficients of the moving averages, where $\bar{I}$ is the sample mean of the absolute difference $\hat{I}$ of the irregular part $\left|\bar{I}_t - \bar{I}_{t-1}\right|$ of the series and $\bar{C}$ is the sample mean of the absolute difference $\left|\bar{C}_t - \bar{C}_{t-1}\right|$ of the trend one cycle part $\hat{C}$ of the series.

Based on the ratio R and the central moving average coefficient, Musgrave gives the formula for the asymmetric moving average coefficient:

$$\varphi_j = \theta_j + \frac{1}{M} \sum_{i=M+1}^{N} \theta_1 + \frac{\left[j - \frac{M+1}{2}\right] D}{1 + \frac{M(M-1)(M+1)}{12}} \sum_{i=M+1}^{N} \left[i - \frac{m+1}{2}\right] \theta_i, -(k-d) \ll j \ll k, \tag{2.2}$$

where $N = 2K + 1$, $M = N - d$, $D = \frac{4}{\pi R^2}$.

We can obtain the asymmetric moving average coefficient through the use of Eq (2.2), and then get the smooth estimation of the missing term.

2.3. The X-12-ARIMA model

X-11 is the core of X-11-ARIMA and X-12-ARIMA [36]. In the process of applying the simple central moving average and Henderson weighted moving average, some fitting values may be missing. The X-11-ARIMA process is to construct the ARIMA model to fill the missing data during the moving average process before establishing the X-11 model. On this basis, the United States Census Bureau strengthened the preprocessing of the sequence and developed the X-12-ARIMA model in 1998 [37]. Figure 1 shows the flow of the X-12-ARIMA process.

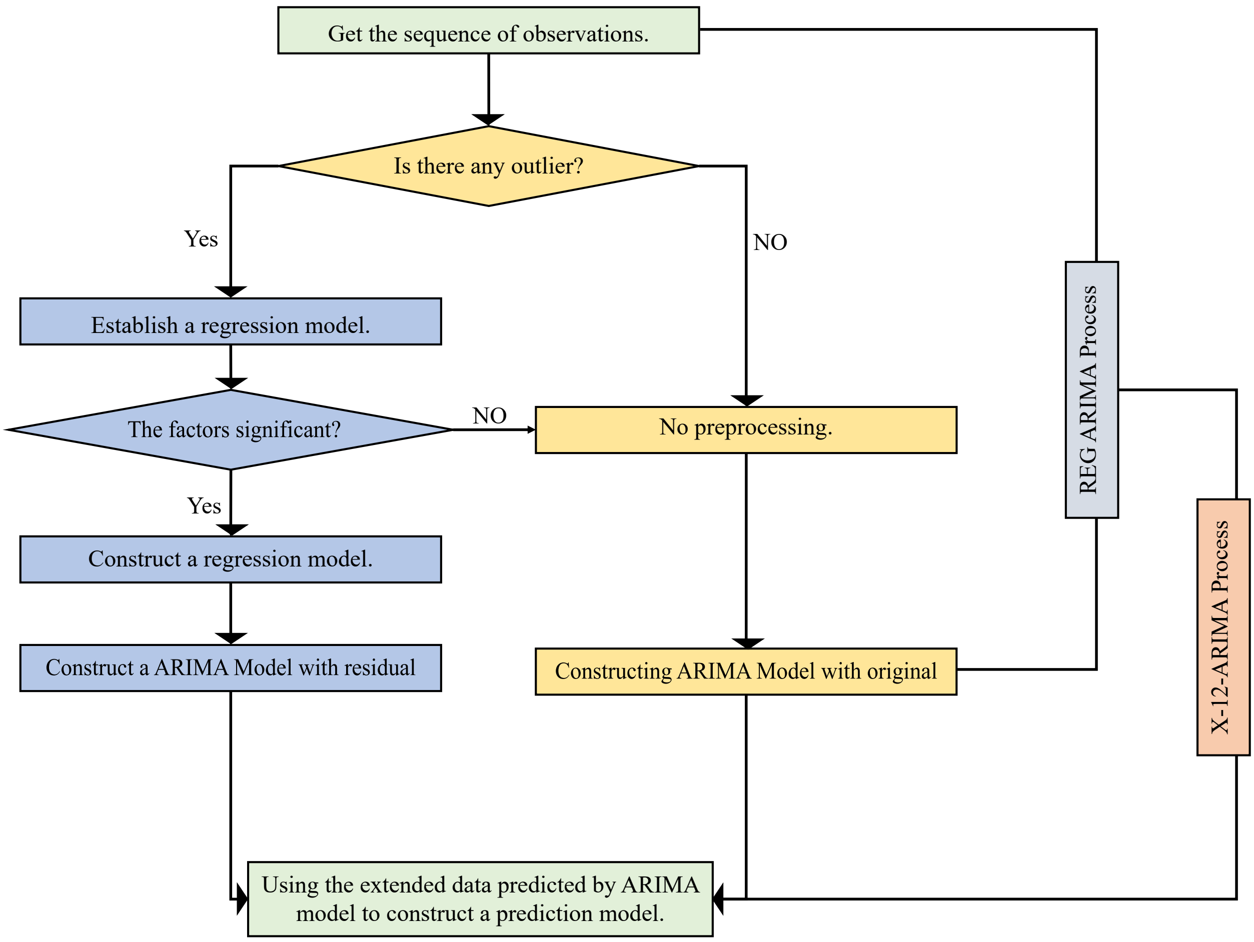

Figure 1. The process of X-12-ARIMA model.

Step 1: Check whether there are any deterministic outliers have an impact on the series values.

The new model strengthens the preprocessing of sequence values by detecting the influence of special factors on the sequence through regression.

Step 2: Construct an ARIMA model according to the fitting results of the regression model.

If the regression equation is significant, construct an ARIMA model with the residual series. Otherwise, construct an ARIMA model with the original sequence.

Step 3: Construct the prediction model by using the expanded data in step 2.

In order to fill the missing data points, the system will use the fitted ARIMA model to predict the data automatically, and then construct the prediction model.

Step 4: Predict the research object with the X-12-ARIMA model.

3. Application of the X-12-ARIMA model

As one of the components of the Yangtze River Delta, the development speed of Anhui Province has remained high in recent years, and the air quality problem is becoming more and more serious. In 2020, Hefei was classified as a new first-tier city. At the same time as economic development, Hefei has contributed to the control of air pollution.

In 2020, the proportion of days with good air quality in Hefei was 84.7%, of which $PM_{2.5}$ exceeded the standard and the air quality did not meet the standard. According to the ranking of urban air quality, Hefei ranks 84th among which is 168 key cities, in the middle level. It ranks 234 among 337 cities at prefecture level and above, which is lower than the national average. Concerned about the standard-exceeding rate of pollutants and the days of primary pollutants, the standard-exceeding rates of O_3, $PM_{2.5}$ and NO_2 were 4.9%, 8.8% and 3.0% respectively. The emissions of sulfides and carbides did not exceed the standard. The pollution days with $PM_{2.5}$, NO and PM as the primary pollutant were 30 days, 5 days and 3 days respectively [38].

The prediction of the Hefei AQI can consider the effectiveness of the existing atmospheric prevention and control policies to some extent, as well as correct the existing policies through forecasting data and models. Here, we selected the monthly data of the Hefei AQI from 2014 to 2021. Taking 30 days before and after the Spring Festival as "special trading days" to analyze whether the policy of fireworks and firecrackers is effective or not.

Figure 2 shows the change of the AQI in Hefei from 2014 to 2021. Figure 3 gives the change of the annual average of the AQI in Hefei during the same period. It can be seen that the AQI roughly shows a downward trend of periodic fluctuations during the study period. Affected by seasonal and diurnal changes, the AQI is generally high in autumn and winter and low in spring and summer. If the policy of banning fireworks during the Spring Festival is useful, then the selected "special trading day" factors will not be significant in the regression analysis. On the contrary, if the influencing factors of the Spring Festival are significant, it shows that the control of fireworks and firecrackers alone can not effectively prevent and control air pollution.

The influential factor of the Chinese New Year is obtained by dividing the number of days in the Spring Festival influential period of each month by the total number of days of the month. Table 1 shows the sequence of influencing factors for the Spring Festival based on the AQI data before and after the Spring Festival.

Based on the influential factors for the Spring Festival in Table 1, the regression model is established by taking the sequence valuea of the influential factors as independent variables and the AQI of the current month as dependent variables. By determining whether the influencing factor is significant in the regression model, we can judge whether the Spring Festival has a significant impact on the AQI.

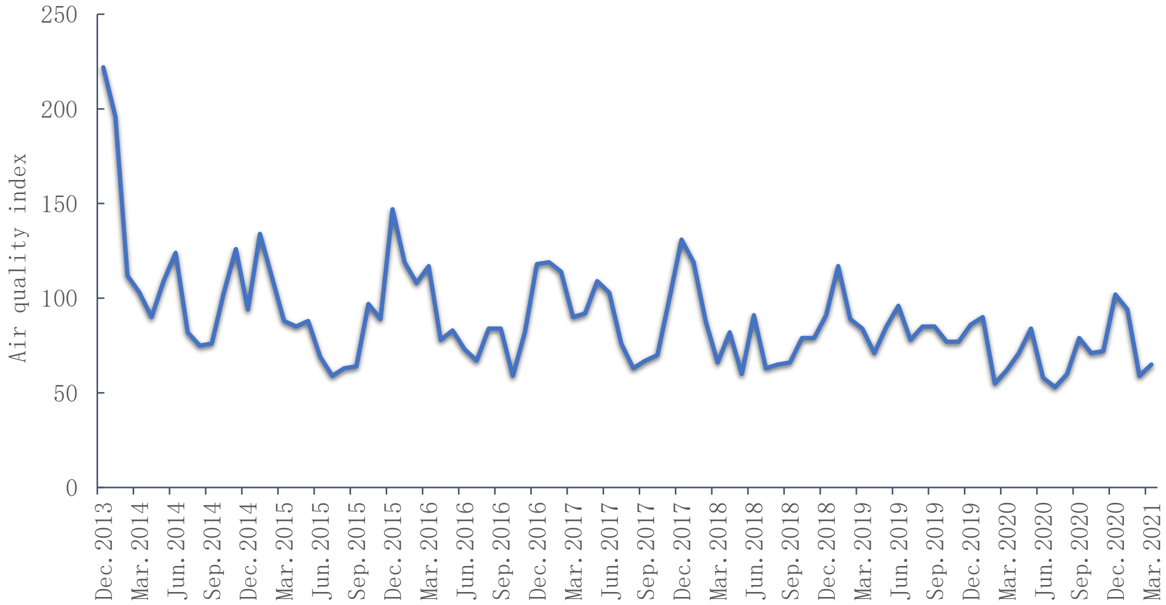

Figure 2. The change of AQI between 2012–2022.

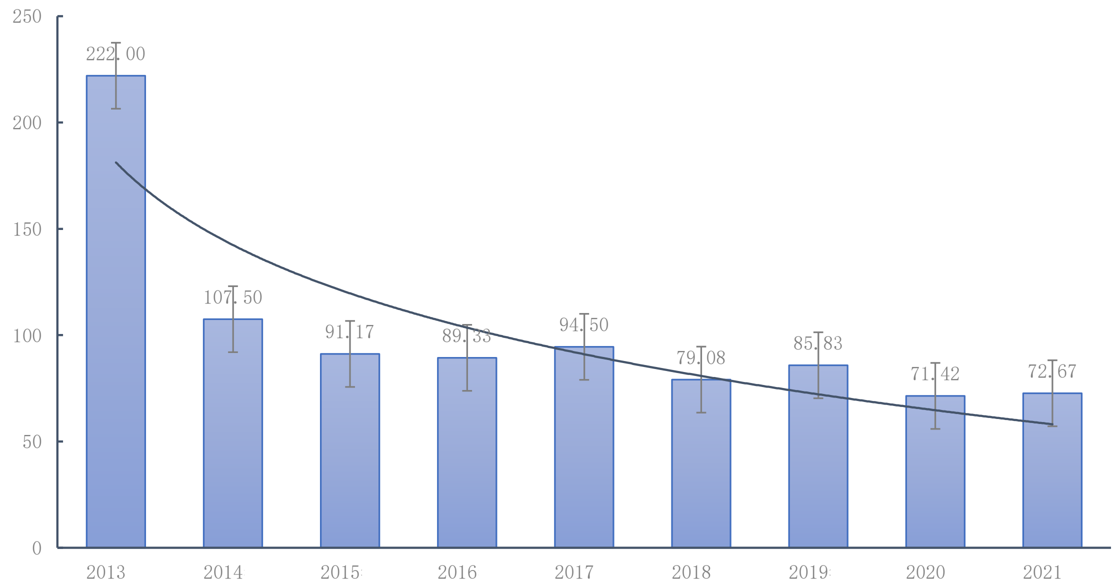

Figure 3. The change in annual average of AQI between 2012–2014.

Table 1. January-March Spring Festival influential factor values during 2014–2020.

Year	Jan.	Feb.	Mar.
2014	16/31	14/28	0
2015	0	25/28	5/31
2016	8/31	22/29	0
2017	0	1	2/31
2018	11/31	19/28	0
2019	0	1	2/31
2020	22/31	8/29	0
2021	4/31	21/29	0

Table 2 shows the fitting values of the influencing factors of the Lunar New Year. It can be seen that the p-value of the influential factor is 0.719, which is significantly higher than the given significance level $\alpha = 0.05$. Therefore, the regression equation cannot be established significantly. In other words, the Spring Festival effect will not affect the AQI sequence of Hefei. There may be changes in the AQI caused by the increase of traffic volume during the Spring Festival, but the impact is not serious. Therefore the Spring Festival effect is not a significant factor affecting the AQI sequence in Hefei. From practical experience, a large number of fireworks and firecrackers will inevitably lead to a serious increase in the AQI.

Thus it can be inferred that the policy of banning fireworks and firecrackers in Hefei has a certain effect on controlling air pollution.

Table 2. The fitting results of the Spring Festival influential factors during 2014–2021.

Models	Unstandardized factor		Standardized factor	t	Significance
	B	Standard error	Beta		
Constant	95.886	17.640	-	5.435	0.000
Index	11.358	30.884	0.98	0.368	0.719

Therefore, before building the X-11 model, we used the original sequence to build the ARIMA model to supplement the sequence values that will be missing in the moving average.

Figure 2 shows the time sequence chart of the AQI, which present that the AQI as a whole has a downward trend of fluctuation , and that the time series is judged to be obviously seasonal. In order to verify the composition of the composite time series, the smoothness analysis and white noise test of the time series were carried out.

The autocorrelation coefficient is not less than twice the standard deviation, which means that the sequence is not smooth. Considering the seasonality and periodicity of the AQI time series, the first-order twelve-step difference was carried out to eliminate the seasonality and periodicity of the series. Then, the white noise test, unit root test and smoothness test were carried out to get the output results of Table 3, Table 4 and Figure 4, respectively.

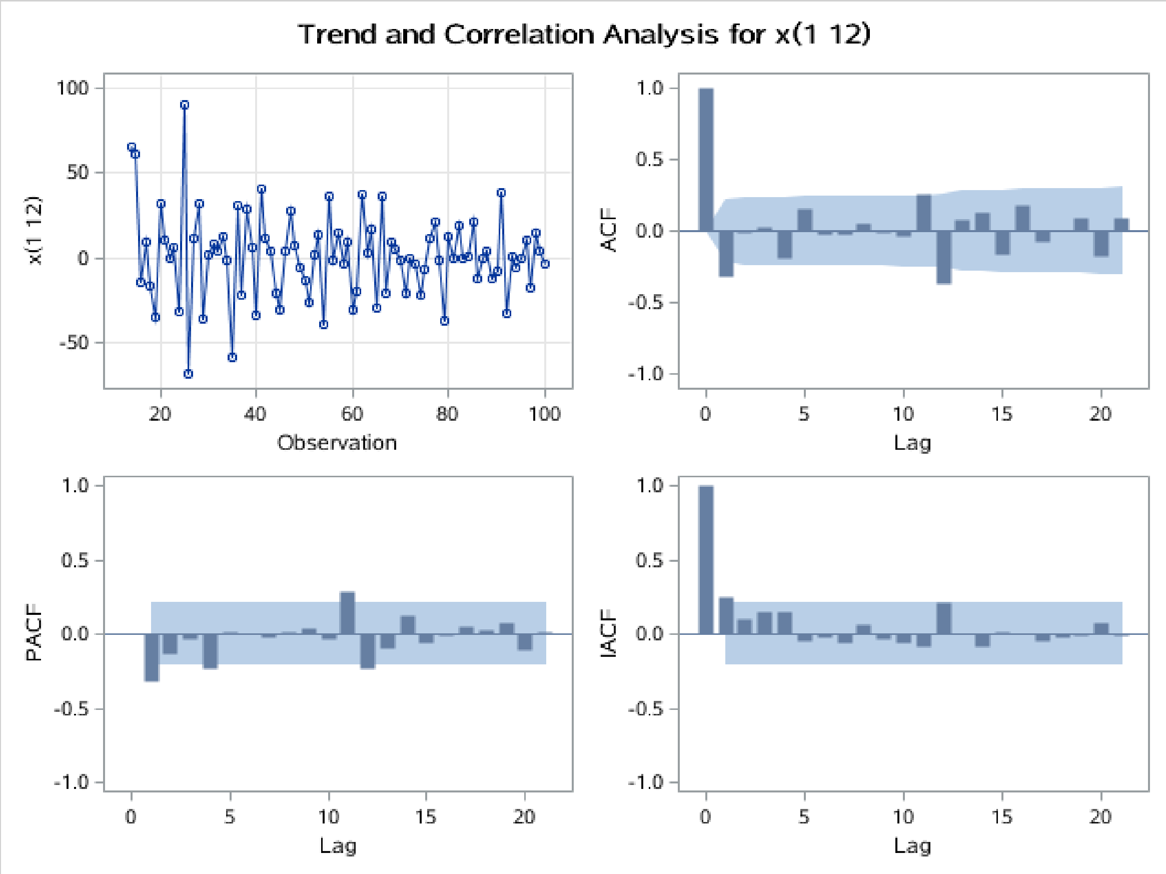

Figure 4. The trend and correlation result for dif_{12} (AQI).

Table 3. The autocorrelation results after the first-order 12-step differencing of the AQI.

To Lag	Chi-Square	DF	Pr >ChiSq	Autocorrelations					
6	15.33	6	0.0178	-0.323	-0.017	0.022	-0.195	0.155	-0.026
12	37.14	12	0.0002	-0.024	0.056	-0.016	-0.032	0.259	-0.375
18	47.02	18	0.0002	0.073	0.135	-0.168	0.185	-0.074	-0.005

Table 4. The unit root test outcome for dif_{12} (AQI).

Type	Lags	Rho	Pr <Rho	Tau	Pr <Tau	F	Pr >F
Zero Mean	0	-113.534	0.0001	-13.45	<.0001		
	1	-170.259	0.0001	-10.50	<.0001		
Single Mean	0	-113.785	0.0001	-13.39	<.0001	89.70	0.0010
	1	-171.045	0.0001	-10.43	<.0001	54.56	0.0010
Trend	0	-114.321	0.0001	-13.36	<.0001	89.36	0.0010
	1	-172.336	0.0001	-10.34	<.0001	53.87	0.0010

The result in Figure 4 is a relatively typical ACF result of smooth time series.

Similarly, the p-values in the unit root test result output from Table 4 were all less than 0.01, and it is judged that the sequence has been classified as a smooth time series. Table 3 shows that the p-values of the differential time series in the white noise test were all less than 0.1, so the series has passed the white noise test and is considered to be a non-white noise series.

After the model was fitted, a residual test was performed to check the fit. If the residual sequence showed pure randomness, it means that the model fit well and there is no need for secondary information extraction of the residual sequence.

Table 5. The residual series test outcome for $\mathrm{dif}_{12}(\mathrm{AQI})$.

To Lag	Chi-Square	DF	Pr >ChiSq	Autocorrelations					
6	5.93	5	0.3127	-0.043	-0.127	-0.048	-0.176	0.112	0.017
12	21.59	11	0.0277	-0.017	0.055	-0.010	0.050	0.174	-0.341
18	26.63	17	0.0637	0.005	0.133	-0.092	0.141	-0.026	0.003
24	29.86	23	0.1536	0.054	-0.147	0.039	0.011	-0.026	-0.029

For this sequence, the residual was analyzed by using autoregression to get Table 5. It can be seen that the p-values of the residual test statistics were all greater than 0.01, and the residual series is considered to be a series of white noise. Therefore, the original model fitting is effective.

Table 6. The fitting value for the ARIMA process.

Parameter	Lag	Estimate	Standard Error	t Value	Pr >|t|
Nonseasonal AR	1	-0.42878	0.10409	-4.12	<.0001
Seasonal AR	12	-0.36934	0.10710	-3.45	0.0009

After getting the ARIMA model, we performed a three-stage and 10-step iterative operation on the supplementary sequence, i.e. the X-11 process. After the above steps, one can obtain the seasonal adjustment model of X-12-ARIMA. The AQI series of Hefei has significant seasonal variation characteristics. It increases significantly in winter every year especially from December to February of the following year (i.e. three months in winter), and decreases obviously in summer, reaching the lowest point in June every year.

After excluding seasonal influences, the trend effect series had a downward trend as a whole. It indicates that the AQI in Hefei decreased significantly from 2014 to 2020. In 2020, the environmental pollution gradually improved and the air quality became better and better during this period.

The fitted values of X-12-ARIMA and its test results are given in Table 6, which shows that the non-seasonal AR1 coefficients and seasonal coefficients are significant. It means that the model is significant. According to the output results, the final fitting model is given by Eq (3.1).

$$(1 - B)(1 - B)^{12} = \frac{\varepsilon_t}{1 + 0.36934B}. \tag{3.1}$$

Table 7. The prediction value of the X-12-ARIMA model.

Month	Data	Forecast	Forecast Error	Month	Data	Forecast	Forecast Error
Sep. 2016	84.000	77.399	6.601	Jun. 2019	96.000	92.246	3.754
Oct. 2016	59.000	115.019	-56.019	Jul. 2019	78.000	70.623	7.377
Nov. 2016	82.000	87.727	-5.727	Aug. 2019	85.000	69.771	15.229
Dec. 2016	118.000	95.666	22.334	Sep. 2019	85.000	82.393	2.607
Jan. 2017	119.000	111.179	7.821	Oct. 2019	77.000	94.934	-17.934
Feb. 2017	114.000	102.467	11.533	Nov. 2019	77.000	96.488	-19.488
Mar. 2017	90.000	105.382	-15.382	Dec. 2019	86.000	101.926	-15.926
Apr. 2017	92.000	74.272	17.728	Jan. 2020	90.000	101.335	-11.335
May. 2017	109.000	84.641	24.359	Feb. 2020	55.000	63.708	-8.708
Jun. 2017	103.000	89.790	13.210	Mar. 2020	62.000	45.676	16.324
Jul. 2017	76.000	92.063	-16.063	Apr. 2020	71.000	52.466	18.534
Aug. 2017	63.000	96.216	-33.216	May. 2020	84.000	65.997	18.003
Sep. 2017	67.000	74.200	-7.200	Jun. 2020	58.000	97.170	-39.170
Oct. 2017	70.000	63.801	6.199	Jul. 2020	53.000	55.618	-2.618
Nov. 2017	100.000	78.490	21.510	Aug. 2020	60.000	50.560	9.440
Dec. 2017	131.000	136.366	-5.366	Sep. 2020	79.000	59.524	19.476
Jan. 2018	119.000	126.361	-7.361	Oct. 2020	71.000	71.352	-0.352
Feb. 2018	88.000	112.156	-24.156	Nov. 2020	72.000	74.689	-2.689
Mar. 2018	66.000	87.579	-21.579	Dec. 2020	102.000	81.773	20.227
Apr. 2018	82.000	56.406	25.594	Jan. 2021	94.000	106.235	-12.235
May. 2018	60.000	80.857	-20.857	Feb. 2021	59.000	70.918	-11.918
Jun. 2018	91.000	67.286	23.714	Mar. 2021	70.000	62.395	7.605
Jul. 2018	63.000	55.625	7.375	Apr. 2021	67.000	66.283	0.717
Aug. 2018	65.000	66.213	-1.213	May. 2021	72.000	81.770	-9.770
Sep. 2018	66.000	66.118	-0.118	Jun. 2021	85.000	64.572	20.428
Oct. 2018	79.000	58.324	20.676	Jul. 2021	48.000	64.257	-16.257
Nov. 2018	79.000	96.900	-17.900	Aug. 2021	56.000	66.669	-10.669
Dec. 2018	91.000	123.866	-32.866	Sep. 2021	70.000	66.908	3.092
Jan. 2019	117.000	93.401	23.599	Oct. 2021	62.000	60.842	1.158
Feb. 2019	89.000	82.190	6.810	Nov. 2021	74.000	62.597	11.403
Mar. 2019	84.000	69.454	14.546	Dec. 2021	87.000	90.567	-3.567
Apr. 2019	71.000	86.595	-15.595	Jan. 2022	94.000	87.558	6.442
May. 2019	85.000	74.899	10.101				

Table 7 shows the prediction and prediction errors made during the X-12-ARIMA process, which translates into percentages indicating that there are still some prediction errors. The disadvantage of the ARIMA model is that it only reflects the short-term autocorrelation of sequences and has some deficiencies in long-term prediction. However, in this case, most of the prediction errors are less than 20%, so the improved ARIMA model based on the seasonally adjusted model has certain applicability and accuracy.

The backward prediction image of the X-12-ARIMA process is shown in Figure 5 which shows that the AQI in Hefei still has a certain fluctuation trend, and that the AQI in Hefei has fluctuated and decreased for more than seven years. It indicates that the air quality control in Hefei has achieved remarkable results in recent years.

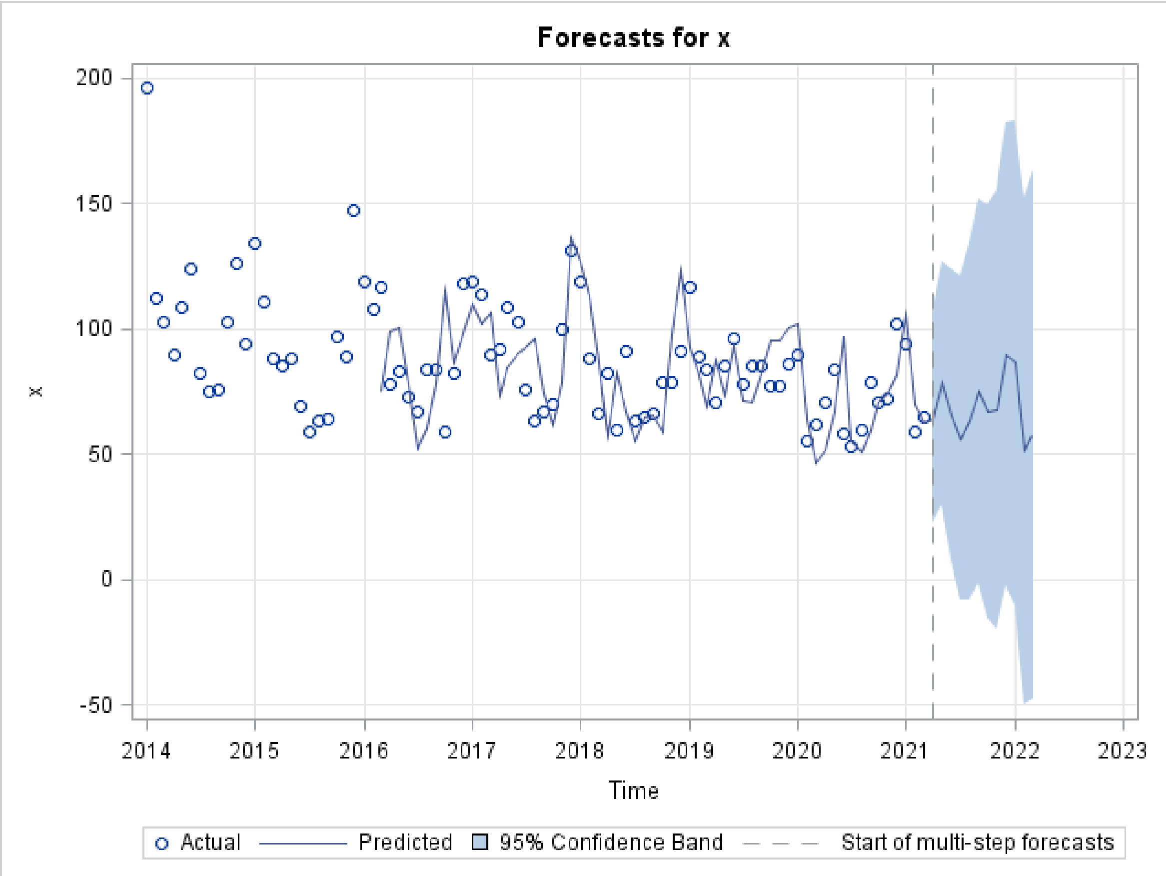

Figure 5. The results of the X-12-ARIMA model prediction.

4. Conclusions and suggestions

4.1. Conclusions

When predicting the AQI, scholars have used different prediction models. Among them, the ARIMA model has a mature theoretical system. To interpret the change of the AQI, scholars take into account the impact of seasonal factors on the time series, but ignore some regular special dates that affect the series value. Such dates are called "special trading days". In order to overcome the ignorance of the special influential factors, the X-12-ARIMA model pays attention to the impact of special trading days when predicting the AQI series. Furthermore, this model also uses three moving average methods to estimate the periodicity, including the special trading days, to predict the changing trend of time series.

We applied the novel model to predict the AQI of Hefei. In 2018, Hefei issued a policy that banned the discharge of fireworks during the Spring Festival. According to the practical experience, the discharge of the fireworks will lead to higher air pollution, making the air quality more severe. We regarded the 30 days before and after the Spring Festival as a special trading days to examine the effectiveness of the fire ban policy in Hefei. Based on the finding, we constructed the improved model, the X-12-ARIMA model, to predict the change of AQI series. The research results can be used to evaluate the air quality prevention policy in Hefei and adjust it on time.

Through the above analysis, we can gain several main conclusions including the following.

(1) The air quality of Hefei shows a trend of fluctuating descent and becomes higher in autumn and winter. In the regression model with the Spring Festival influential factors as the independent variable and the monthly AQI as the dependent variable, the independent variable is not significant. It means that the impact of social behavior on the AQI during the Spring Festival is not obvious. The reason is that Hefei is located in southern area and does not belong to the city that provides heater. Hefei has implemented the control policy on fireworks during recent years, the results of analysis imply that the policy of controlling the emission of pollutants is effective.

(2) The fitting results of the X-12-ARIMA model show that the AQI of Hefei has an obvious descending trend during the study period. It indicates that the prevention and control of air pollution in Hefei is valuable. Hefei has made contributions to energy conservation. It can be seen that this city will still maintain the reduction of pollutant emission while developing the economically.

(3) The AQI is an important indicator is that closely related to human health. The prediction result shows that the AQI of Hefei will continue to show a trend of fluctuating decline in the near future. Economic development is always accompanied with technological growth. Hefei insists on innovation-driven development and reducing primary and secondary pollutants, and it will still drive the improvement of the air quality.

4.2. Suggestions

Based on the analysis above, we present related suggestions and solutions to implement below.

(1) The increase in air pollution in autumn and winter indicates that the pollutant emissions are higher during these seasons. Therefore, Hefei needs to strengthen the control of the pollutant discharge in autumn and winter. For example, it can run heating equipment with new types of energy rather than energy-intensive sources.

(2) It is clear to see that the prevention and control of air pollution in Hefei is appropriate according to the results of the discussion. Hefei can not only ensure its economic development, but also control the degree of air pollution. We suppose that Hefei can continue to implement the existing policies of air pollution prevention and control.

(3) It is necessary to speed up the transformation of the industrial and energy structure. The government should improve its innovation ability, develops the clean energy vigorously and strengthens the new technologies for energy saving and emission reduction.

(4) Anhui Province can take the development mode of Hefei as a reference for green development since it has demonstrated excellent performance. It can ensure the common development of the whole province.

Use of AI tools declaration

The authors declare they have not used artificial intelligence tools in the creation of this article.

Source of data

The data are from the China Air Quality Online Monitoring and Analysis platform (https://www.aqistudy.cn/) and Atmospheric Composition Analysis Group (https://sites.wustl.edu/acag/datasets/).

Acknowledgments

This work was supported in part by Anhui Provincial Key Education and Teaching Research Project (Grant 2022jyxm304) and Natural Science Fund of Education Department of Anhui Province (Grant KJ2020A0478).

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. A. Peters, C. A. Pope III, Cardiopulmonary mortality and air pollution, *Lancet*, **360** (2002), 1184–1185. https://doi.org/10.1016/S0140-6736(02)11289-X
2. C. A. Pope III, R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, et. al., Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, *JAMA*, **287** (2002), 1132–1141. https://doi.org/10.1001/jama.287.9.1132
3. P. Gallagher, W. Lazarus, H. Shapouri, R. Conway, F. Bachewe, A. Fischer, Cardiovascular disease—risk benefits of clean fuel technology and policy: A statistical analysis, *Energ. Policy*, **38** (2010), 1210–1222. https://doi.org/10.1016/j.enpol.2009.11.013
4. D. C, Shin, *Hazardous air pollutants: Case studies from Asia*, 1 Eds., Boca Raton: Press, 2016. https://doi.org/10.1201/b19829
5. C. Mora, D. Spirandelli, E. C. Franklin, J. Lynham, M. B. Kantar, W. Miles, et. al., Broad threat to humanity fromcumulative climate hazards intensifiedby greenhouse gas emissions, *Nat. Clim. Change*, **8** (2018), 1062–1071. https://doi.org/10.1038/s41558-018-0315-6
6. H. X. Zhang, X. F. Cheng, R. H. Chen, Study on Spatio-temporal Distribution characteristics and key influencing factors of PM2.5 in Anhui Provinceg, *Acta Sci. Circumstantiae*, **38** (2018), 1080–1089.
7. Q. S. Zhu, C. W. Xie, J. B. Liu, On the impact of the digital economy on urban resilience based on a spatial Durbin model, *AIMS Mathematics*, **8** (2023), 12239–12256. https://doi.org/10.3934/math.202361
8. X. Dai, G. Song, X. Jiang, X. Yu, D. Yu, Impact of the new crown pneumonia outbreak on air quality in Xianyang City, *China Environ. Sci.*, **41** (2021), 3106–3114. https://doi.org/10.19674/j.cnki.issn1000-6923.20210331.004

9. H. Liu, W. Xu, M. Wei, P. Xu, M. Li, M. Zhang, Simulation of the impact of epidemic control on air quality in Shandong Province in early 2020, *Environ. Sci.*, **42** (2021), 1215–1227. https://doi.org/10.13227/j.hjkx.202007246
10. C. Fang, lmportant progress and prospects of China's urbanization and urban agglomeration in thepast 40 years of reform and opening-up, *Econ. Geogr.*, **38** (2018), 38: 1–9. https://doi.org/10.15957/j.cnki.jjdl.2018.09.001
11. Q. Wan, X. Luo, F. Pan, G. Jin, Spatio-temporal evolution and convergence trend of air quality in urban agglomeration in China, *Geoscience*, **2242** (2022), 1943–1953. https://doi.org/10.13249/j.cnki.sgs.2022.11.009
12. J. B. Liu, X. B. Peng, J. Zhao, Analyzing the spatial association of household consumption carbon emission structure based on social network, *J. Comb. Optim.*, **79** (2023), 45–79. https://doi.org/10.1007/s10878-023-01004-x
13. J. B. Liu, B. Y. Zhao, Study on environmental efficiency of Anhui province based on SBM-DEA model and fractal theory, *Fractals*, **31**, (2023), 2340072. https://doi.org/10.1007/s10878-023-01004-x
14. N. Huang, S. Zhu, The report card of the ten years of industrial development in Hefei, Available from: http://hfdx.hfzhi.com/index/index/index/id/4#magazine/page66-page67.
15. W. Huang, D. Li, Y. Huang, A spatio-temporal hybrid prediction model for air quality, *Comput. Appl.*, **40** (2020), 3385–3392. https://doi.org/10.11772/j.issn.1001-9081.2020040471
16. Y. Zhou, Prediction of air quality in Nanjing based on ARIMA and long-term and short-term memory model, Master thesis, Nanjing Audit University, 2021. https://doi.org/10.27835/d.cnki.gnjsj.2021.000124
17. S. Gautam, A. Yadav, C. J. Tsai, P. Kumar, A review on recent progress in observations, sources, classification and regulations of $PM_{2.5}$ in Asian environments, *Environ. Sci. Pollut. Res.*, **23** (2016), 2116–2117. https://doi.org/10.1007/s11356-016-7515-2
18. G. E. Kulkarni, A. A. Muley, N. K. Deshmukh, P. U. Bhalchandra, Modeling Eart Autoregressive integrated moving average time series model for forecasting air pollution in Nanded city, *Model. Earth Syst. Environ.*, **4** (2018), 1435–1444. https://doi.org/10.1007/s40808-018-0493-2
19. V. Naveen, N. Anu, Time Series Analysis to Forecast Air Quality Indices in Thiruvananthapuram District, Kerala, India, *Int. J. Eng. Res. Appl.*, **7** (2017), 66–84. https://doi.org/10.9790/9622-0706036684
20. H. Zhang, S. Zhang, P. Wang, Y. Qin, H. Wang, Forecasting of particulate matter time series using wavelet analysis and wavelet-ARMA/ARIMA model in Taiyuan, *J. Air Waste Manage. Assoc.*, **67** (2017), 776–788. https://doi.org/10.1080/10962247.2017.1292968
21. E. Aladağ, Forecasting of particulate matter with a hybrid ARIMA model based on wavelet transformation and seasonal adjustment, *Urban Clim.*, **39** (2021), 100930. https://doi.org/10.1016/j.uclim.2021.100930
22. H. Gong, H. Wang, W. Liang, X. Ma, L. Yang, F. Guo, Factor analysis of haze formation in Beijing-Tianjin-Hebei region, *China Environ. Protec. Ind.*, **269** (2020), 34–39. https://doi.org/10.3969/j.issn.1006-5377.2020.11.005
23. X. Huang, T. Shao, J. Zhao, J. Cao, D. Yue, X. Lu, Temporal and spatial distribution characteristics and influencing factors of air quality in the Yangtze River economic belt, *China Environ. Sci.*, **40** (2020), 874–884. https://doi.org/10.19674/j.cnki.issn1000-6923.2020.0149

24. L. Jiang, H. Zhou , L. Bai, Z. Chen, Analysis of socio-economic influencing factors of Air quality Index (AQI) – from the perspective of exponential attenuation effect, *J. Environ. Sci.*, **38** (2018), 390–398. https://doi.org/10.13671/j.hjkxxb.2017.0181
25. G. B. Christopher, H. H. Scott, C. M. Brian, Comparison of X-12-ARIMA Trading Day and Holiday Regressors with Country Specific Regressors, *J. Off. Stat.*, **26** (2010), 371–394.
26. W. M. Persons, An Index of General Business Conditions, *Rev. Econ. Stat.*, **9** (1919), 20–29. https://doi.org/10.2307/1928562
27. C. Chatfield, D. Prothrto, Box-Jenkins seasonal forecasting: Problems in a casestudy, *J. Roy. Stat. Soc.*, **136** (1973), 295–336. https://doi.org/10.2307/2344994
28. S. Markidakis, A survey of time series, *Int. Stat. Rev.*, **44** (1976), 29–70. https://doi.org/10.2307/1402964
29. J. Tang, X. Zhong, J. Liu, T. Li, Short-term prediction of rail transit passenger flow based on time series seasonal classification model, *J. Chongqing Jiaotong Univ.*, **40** (2021), 31–38. https://doi.org/10.3969/j.issn.1674-0696.2021.07.05
30. L. Dominique, B. Quennevill, *Seasonal Adjustment with the X-11 Method*, New York: Springer, 2001. https://doi.org/10.1007/978-1-4613-0175-2
31. W. Fan, L. Zhang, G. Shi, Summary and comparison of seasonal adjustment methods, *Stat. Res.*, **2** (2006), 70–73. https://doi.org/10.19343/j.cnki.11-1302/c.2006.02.018
32. W. P. Cleveland, G. C. Tiao, Decomposition of Seasonal Time Series: A Model for the Census X-11 Program, *J. Am. Stat. Assoc.*, **71** (1974), 581–587. https://doi.org/10.6092/issn.1973-2201/3597
33. J. Shiskin, The Census Bureau Seasonal Adjustment Program, *Bus. Econ.*, **4** (1969), 71–73.
34. X. Liu, The application and enlightenment of seasonal adjustment method in western countries, *Jiangsu Stat.*, **11** (1999), 32–34.
35. R. J. Hyndman, Moving Averages, *Int. Encycl. Stat. Sci.*, **eds** (2011), 866—869. https://doi.org/10.1007/978-3-642-04898-2_380
36. S. Liu, On the Seasonal Adjustment of Time Series Data, the Derivation of Quarterly Change Rate and Annual Change Rate and the Annualization Method, *Res. World Econ. Stat.*, **1** (2003), 15–21.
37. D. F. Findley, B. C. Monsell, W. R. Bell, M. C. Otto, B. Chen, New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program, *J. Bus. Econ. Stat.*, **16** (1998), 127–152. https://doi.org/10.1080/07350015.1998.10524743
38. L. He, M. Zhou, Y. Zhu, X. Meng, D. Hu, Q. Fu, et. al., Study on the changing trend of air quality in Hefei from 2001–2020, *China Environ. Monit.*, **38** (2022), 65–73. https://doi.org/10.19316/j.issn.1002-6002.2022.04.08