The impact of air quality on its Baidu index: grey model analysis

Leping Tu

Hebei University of Engineering

Yan Chen (✉ 964129856@qq.com)

Hebei University of engineering

Research Article

Keywords: air quality, Baidu index, grey multivariate convolution model, particle swarm optimization

Posted Date: February 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-202216/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

The impact of air quality on its Baidu index: grey model analysis

Leping Tu, Yan Chen*

School of Management Engineering and Business, Hebei University of Engineering, Handan, 056038, China

Corresponding Author: E-mail address: 964129856@qq.com

Abstract: To investigate the relationship between air quality and its Baidu index, we collect the annual Baidu index of air pollution hazards, causes and responses. Grey correlation analysis, particle swarm optimization and grey multivariate convolution model are used to simulate and forecast the comprehensive air quality index. The result shows that the excessive growth of the comprehensive air quality index will lead to an increase in the corresponding Baidu index. The number of search for the causes of air quality has the closest link with the comprehensive air quality index. Strengthening the awareness of public about air pollution is conducive to the improvement of air quality. The result provides a reference for relevant departments to prevent and control air pollution.

Keywords: air quality; Baidu index; grey multivariate convolution model; particle swarm optimization

1. Introduction

Air quality prediction is the basis of environmental protection work and plays a crucial role in environmental protection [1]. It can provide scientific data for environmental management, environmental assessment, and environmental planning [2]. Air quality level is an important standard for measuring the quality of the environment, which directly affects people's health [3, 4]. Therefore, it is particularly important to strengthen the prediction of air pollution and timely grasp the air quality status. Only by strengthening the prevention and control of air pollution can we improve air quality further [5]. After

strengthening the prediction of air pollution, we can understand the development trends of air pollution. The prediction results combined with its current status can provide relevant data support for air pollution treatment [6]. Continuous optimization and improvement of air pollution control programs can better promote the effectiveness of air pollution control. In recent years, with the concept of environmental protection gradually gaining popular support, the country's attention to environmental protection has been increasing. Especially in the area of air pollution prediction and control, efforts are constantly improving. However, at present, air pollution in some cities in China is more serious, and air quality prediction and improvement methods need to be further improved [7].

The widespread use of big data has brought good news to air prediction [8, 9]. The traditional data for air quality prediction are mostly from statistical reports of the government and environmental protection departments. Since the completion of the report requires a certain amount of time, the collection and release of these data often lack timeliness and cannot directly reflect the current problem. The lack of data has caused huge obstacles to prediction [10]. No matter how advanced the prediction methods and tools used by researchers, the data source is still the bottleneck of air prediction. Therefore, the use of data that more timely and directly reflects air quality will greatly improve the prediction effect. In recent years, network data from search engines has been widely used in prediction research in various fields. Tourism forecasting, financial forecasting and hospitalization forecasting based on network search data are all have higher accuracy [11, 12]. The instantaneity of network data can make up for the lag of traditional forecasting data and has stronger timeliness. Therefore, the characteristics of web search data determine its applicability in the prediction of social and economic activities. By extracting rules from a large amount of search data and analyzing data change trends, it can provide effective methods for real problems [13]. The increasing popularity of Internet technology in China has made the spread of information no longer limited by time

and space. The platform has also become an important source of information for the public [14]. Therefore, web search data can be used to mine the public's perception and response to air quality. A proper use of network data is conducive to reliable air quality prediction.

At present, most air quality forecasts are short-term with a period of hours and days. For example, the Light Gradient Boosting Machine model is used to predict the $PM_{2.5}$ concentration in Beijing in the next 24 hours [15]. Improved model based on neural network is used to predict $PM_{2.5}$ pollution at air quality testing stations for 48 hours [16]. Using hourly $PM_{2.5}$ concentration data collected from Beijing's 1,233 air quality monitoring stations, the air quality is predicted [17]. Composed of complexity analysis, data preprocessing, and optimized prediction modules, an analysis and prediction system is used to predict the hourly AQI sequence of eight cities in China [18]. Therefore, this paper will perform an annual air quality forecast in order to propose longer-term air quality improvement methods. Existing annual data are limited, and traditional statistical methods are inappropriate in this case. The grey model will be taken to deal with this problem.

Grey prediction theory is put forward by Professor Deng to deal with the problem of uncertain data or small sample size [19]. Grey forecasting models have attracted much attention because of their remarkable effect [20-22]. In the case study of e-waste data in Washington, the nonlinear grey Bernoulli model with convolutional integral was optimized by particle swarm optimization algorithm, which improved the accuracy of the model [23]. The unbiased fractional discrete multivariate grey model is used to predict the power consumption [24]. The grey model is optimized by changing the fractional order in order to better predict electricity usage [25]. The lion ant colony optimization algorithm is designed to determine the optimal accumulation coefficient to further improve the predictive accuracy of the grey model [26]. To discuss the relationship between CO_2 emission and economic growth, unequal gap grey Verhulst model was

derived [27]. Many scholars also applied grey system theory to the study of air quality [28, 29]. However, few people apply the grey model to study the relationship between web search information and air quality. Compared to AQI, Comprehensive air quality index (CAI) is more convincing in assessing air quality [30]. In this paper, the relationship between Baidu index and CAI is analyzed by grey multivariable convolution model.

The rest of this article is organized as follows. Section 2 gives the research area and data sources. Section 3 introduces research methods. Section 4 analyzes the relationship between Baidu index and CAI in Beijing. The conclusion and implication are drawn in Section 5.

2. Research area and data source

Beijing is China's capital and the center of China's political, cultural and economic development (Fig.1). However, due to the rapid economic development and the rapid increase in population, coupled with the special terrain and atmospheric characteristics of the Beijing-Tianjin-Hebei region, Beijing is encountering a serious air pollution problem. This problem not only endangers public health and exacerbates the deterioration of the natural environment, but also affects the normal economic production and social order. Therefore, it is particularly significant to predict and improve air quality in Beijing.

Fig. 1. Location of Beijing

Internet search information is generated from the public's spontaneous Internet search behavior, which can directly reflect the public's intentions. It can reflect the changes in air quality precisely with the characteristics of real-time and public scale. Facing air pollution, the public often obtains knowledge from others to solve problems due to the lack of relevant knowledge and information. Hence, the public's attention to the problem and the need to understand the relevant information will give rise to corresponding information search and query behavior. In Internet users' search for air quality issues, specific search terms are used to obtain information about air quality, such as "what can be done to reduce air pollution." To form a combination of network search keywords that can be used in specific prediction models. Common sources of web information for web users include search engines, portals, forums, microblogs, and other social software. The Baidu index is a data reference platform provided by Baidu. The search index provided in the platform is calculated based on the number of keywords searched by netizens. Therefore, Baidu index is the research data sources.

People perceive air quality problems when their bodies are affected by air pollution. According to the theory of protection motivation [31], perceived risk determines the willingness to respond. Therefore, we

classify the network search data from the hazards, causes and responses of air pollution. First of all, selecting "bronchitis, smog, environmental protection" as the benchmark keywords by consulting relevant experts and considering the availability of keywords. Then, we expand the keywords with the hot word recommendation feature of search engines. The keyword list is shown in Table 1.

Table 1 The keywords of Baidu index

Hazards (H-BI)	Causes (C-BI)	Responses (R-BI)
myopia	UV	new energy (NE)
acid rain	sulfur dioxide	surroundings protection (SP)
chest tightness	fog	environmental protection (EP)
pulmonary Edema	haze	Environmental Protection Agency (EPA)
difficulty breathing	PM ₁₀	protect environment (PE)
inflammation	CO	air quality (AQ)
lung cancer	smog	surroundings (SUR)
	PM _{2.5}	air purification(AP)

The data in this paper are from China Environmental Monitoring Station (<http://www.cnemc.cn/>) and Baidu index official website (<http://index.baidu.com/>). The available Baidu index are limited, we can only collect available Baidu index from 2013 to 2018. Due to the limited annual data available, traditional statistical methods are not appropriate. This article adopts the grey model to solve this problem.

3 Research methods

3.1 Grey multivariable convolution model

Because of its good predictive effect, GMC(1,N) model is widely used in predictive analysis in various fields. In order to make better predictions, fractional order accumulation is used for optimization.

Its definition is as follows.

Definition The original nonnegative sequence is

$$X_i^{(0)} = \{x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(m)\}, i = 1, 2, \dots, N.$$

By using $x^{(r)}(k) = \sum_{j=1}^k C_{k-j+r-1}^{k-j} x^{(0)}(j), 0 < r < 1$, the r-order accumulation generation operation sequence is

$$X_i^{(r)} = \{x_i^{(r)}(1), x_i^{(r)}(2), \dots, x_i^{(r)}(m)\},$$

Where $C_{r-1}^0 = 1, C_k^{k+1} = 0, C_{k-j+r-1}^{k-j} = \frac{(k-j+r-1)(k-j+r-2) \cdots (r+1)r}{(k-j)!}$

The whitened equation of GMC(1,N) model is written as

$$\frac{dx_1^{(1)}(t)}{dt} + b_1 x_1^{(1)}(t) = b_2 x_2^{(1)}(t) + b_3 x_3^{(1)}(t) + \cdots + b_N x_N^{(1)}(t) + u \quad (1)$$

The grey derivative is noted as $\frac{dx_1^{(1)}(t)}{dt} = x_1^{(1)}(t+1) - x_1^{(1)}(t)$. $b_1, b_2, \dots, b_N$ and u are parameters. They are calculated by the least-squares as $[\hat{b}_1, \hat{b}_2, \dots, \hat{b}_N, \hat{u}]^T = (B^T B)^{-1} B^T Y$.

Where,

$$Y = \begin{bmatrix} x_1^{(1)}(2) - x_1^{(1)}(1) \\ x_1^{(1)}(3) - x_1^{(1)}(2) \\ \vdots \\ x_1^{(1)}(m) - x_1^{(1)}(m-1) \end{bmatrix}$$

$$B = \begin{bmatrix} -\frac{x_1^{(1)}(1)+x_1^{(1)}(2)}{2} & \frac{x_2^{(1)}(1)+x_2^{(1)}(2)}{2} & \dots & \frac{x_N^{(1)}x_N^{(1)}(1)+x_N^{(1)}(2)}{2} & 1 \\ -\frac{x_1^{(1)}(2)+x_1^{(1)}(3)}{2} & \frac{x_2^{(1)}(2)+x_2^{(1)}(3)}{2} & \dots & \frac{x_N^{(1)}(2)+x_N^{(1)}(3)}{2} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{x_1^{(1)}(m-1)+x_1^{(1)}(m)}{2} & \frac{x_2^{(1)}(m-1)+x_2^{(1)}(m)}{2} & \dots & \frac{x_N^{(1)}(m-1)+x_N^{(1)}(m)}{2} & 1 \end{bmatrix}.$$

Then, the approximate time-response function of GMC(1,N) is

$$\hat{x}_1^{(1)}(t) = x_1^{(0)}(1)e^{-b_1(t-1)} + \sum_{\tau=2}^t \left\{ e^{-b_1(t-\tau+0.5)} \frac{f(\tau) + f(\tau-1)}{2} \right\}.$$

Where $f(t) = b_2x_2^{(1)}(t) + b_3x_3^{(1)}(t) + \dots + b_Nx_N^{(1)}(t) + u$. By inverse cumulative generation operator processing,

the predicted value is

$$\hat{x}_1^{(0)}(1) = x_1^{(0)}(1), \hat{x}_1^{(0)}(k) = \hat{x}_1^{(1)}(k) - \lambda \hat{x}_1^{(1)}(k-1), k = 2, 3, \dots, m.$$

In this paper, the mean absolute percentage error (MAPE) is used for evaluating the models, which is calculated as:

$$\text{MAPE} = \frac{\sum_{t=1}^n |(A_t - F_t) / A_t|}{n}$$

Where A_t is the actual value, F_t is the simulated values, and n is the number of years. The smaller the MAPE, the closer to the actual value and the more accurate is the forecasting model.

3.2 Particle swarm optimization

The particle swarm optimization (PSO) algorithm is a commonly parameter optimization algorithm, which iteratively updates the initialized particles until it reaches the optimal solution or the number of iterations is exhausted. It is used to optimize the fractional order to achieve the desired result in this paper.

Step 1: The position and velocity values of the first generation particles are randomly generated.

Step 2: Storing the calculated position and fitness of each particle in $pbest(Q_i^b)$. The position and fitness of the best-fit individuals in all $pbest$ are stored in $gbest(Q_g^b)$. $pbest, gbest$ are the best position experienced by the individual and the global best position.

Step 3: Update the speed and displacement of particles as follow:

$$V_i^k = \omega_k V_i^{k-1} + c_1 r_1 (Q_i^b - Q_i^{k-1}) + c_2 r_2 (Q_g^b - Q_i^{k-1}), i = 1, 2, \dots, n.$$

$$Q_i^k = Q_i^{k-1} + V_i^k, i = 1, 2, \dots, n.$$

Where ω is inertia factor, k and i are the numbers of generation and particles, c_1, c_2 is learn factor,

r_1, r_2 is random number.

Step 4: All particles are compared with the particle of $pbest$. The better is used as the current $pbest$.

Each $pbest$ is compared with the last iteration $gbest$, the better is used as the current optimal $gbest$.

Step 5: If the optimal precision is reached or the number of iterations is exhausted, the operation is stopped. Otherwise, cycle to step 2.

The relevant parameters of PSO in this paper are shown in Table 2.

Table 2 Parameter values of PSO

Number of particles	Number of generations	ω	$c_1 = c_2$	$P_{\max}$	$P_{\min}$	r_1	r_2
30	50	0.9	2	0.1	1.0	0.6	0.3

4. Empirical study

We will discuss the relationship between Baidu index and CAI from three aspects: hazards, cause and

response. Correlation analysis was performed on Baidu index and CAI by grey correlation analysis. These data are summed with correlation degree as the weight to get the model input data sequence. The GMC(1,N) and PSO algorithm were used to obtain prediction results and MAPE values. The specific process is as follows:

Fig. 2 Flow chart of the study process

4.1 The relationship between H-BI and CAI

Table 3 shows the value of CAI and H-BI from 2013 to 2018. CAI is the reference sequence. H-BI is the comparison sequence. The correlation degrees are calculated by grey correlation analysis are shown in Table 4.

Table 3 the value of CAI and H-BI from 2013 to 2018

Year	2013	2014	2015	2016	2017	2018
CAI	5.26	7.56	6.78	6.08	5.57	5.04
myopia	126	166	171	170	167	183
acid rain	101	141	141	141	139	146
chest tightness	140	190	199	194	202	205
pulmonary Edema	93	127	131	132	132	139
difficulty breathing	99	146	132	133	150	154
inflammation	102	143	160	170	154	164
lung cancer	294	386	409	424	429	495

Table 4 Grey relational degree between H-BI and CAI

myopia	acid rain	chest tightness	pulmonary Edema	difficulty breathing	inflammation	lung cancer
0.7035	0.6991	0.6907	0.6724	0.6491	0.6298	0.6284

Then the weight of each factor of H-BI is determined by the grey correlation analysis. Finally, a new number sequence is formed. The sequence is named H-value. H-value and CAI are given in Table 5. Then we can obtain the simulation value of GMC(1,2) model. To make the simulation more convincing, the simulation is divided into two parts: fitting and verification. During the fitting process, the H-value from 2013 to 2016 is used to fit the value of CAI from 2013 to 2016. In order to make the MAPE of fitting as small as possible, the accumulation operator λ of the GMC(1,2) model is determined by the PSO algorithm. The verification process is used to verify the predictive effect of the GMC(1,2) model. During the verification process, the value of CAI from 2017 to 2018 is obtained by putting H-value from 2017 to 2018 into the GMC(1,2) model.

Table 5 the value of CAI and H-value from 2013 to 2018

Year	2013	2014	2015	2016	2017	2018
CAI	5.26	7.56	6.78	6.08	5.57	5.04
H-value	135.19	183.90	190.00	192.76	194.07	209.75

$$x_1^{(0)} = (5.26, 7.56, 6.78, 6.08)$$

$$x_2^{(0)} = (135.19, 183.90, 190.00, 192.76)$$

When $\lambda = 0.10$

$$Y = \begin{bmatrix} 7.044155978 \\ 6.052523107 \\ 5.27638955 \end{bmatrix} \quad B = \begin{bmatrix} -8.78458 & 221.5516 \\ -15.3329 & 392.333 \\ -20.9974 & 560.6383 \end{bmatrix}$$

$$\begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ u \end{bmatrix} = \begin{bmatrix} 0.254927 \\ 0.003968 \\ 8.404394 \end{bmatrix}.$$

By the ordinary least square method, the simulated value and error of fitting and verification are shown in Table 6. The MAPE of fitting and verification are 0.501% and 0.809%. The results show that the model can predict CAI well.

Table 6 The simulated value and error of GMC(1,2)

	Year	Actual CAI	Simulated value	MAPE
fitting	2013	5.26	5.26	0.501%
	2014	7.56	7.50	
	2015	6.78	6.73	
	2016	6.08	6.05	
verification	2017	5.57	5.48	0.809%
	2018	5.04	5.04	

The value of CAI from 2019 to 2023 is predicted under the assumption of increase rate of H-value.

The increase rates of H-value are set as -10%, -5%, 5%, 10% and 20% respectively. Then H-value are put into the GMC(1,2) model to predict the CAI of the corresponding year. The results are shown in Table 7.

Table 7 CAI at different growth rates for H-value

Year	H-value increase rate				
	-10%	-5%	5%	10%	20%
2019	4.68	4.69	4.73	4.75	4.79
2020	4.31	4.37	4.51	4.59	4.74
2021	3.95	4.08	4.38	4.54	4.89
2022	3.60	3.82	4.31	4.59	5.23
2023	3.28	3.57	4.29	4.73	5.78

As shown in Table 7, CAI from 2019 to 2023 shows different trends under five different H-value growth rates. It can be seen more intuitively from Fig. 3. CAI increases when H-value growth rate increases. CAI decreases when H-value growth rate decreases. Moreover, CAI increases or decreases faster when the added value becomes larger or smaller. It indicates that H-value has a significant effect on the value of CAI.

Fig. 3. CAI for different growth rates of H-BI

4.2 The relationship between C-BI and CAI

Table 8 shows the value of CAI and C-BI from 2013 to 2018. The prediction process is same as the Part 4.1. The results are shown in Table 9.

Table 8 the value of CAI and C-BI from 2013 to 2018

Year	2013	2014	2015	2016	2017	2018
CAI	5.26	7.56	6.78	6.08	5.57	5.04
UV	112	155	157	160	168	178
sulfur dioxide	122	165	184	192	190	193
fog	124	158	165	162	158	162
haze	227	227	222	206	166	139
PM ₁₀	109	148	154	149	161	182

CO	133	178	198	204	207	235
smog	930	868	1119	1007	987	802
PM _{2.5}	1561	3286	2721	2768	2631	2071

Table 9 CAI at different growth rates for C-BI value

Year	C-BI value increase rate				
	-10%	-5%	5%	10%	20%
2019	4.44	4.48	4.55	4.59	4.67
2020	3.96	4.09	4.36	4.50	4.79
2021	3.53	3.77	4.31	4.61	5.25
2022	3.15	3.51	4.36	4.85	6.00
2023	2.82	3.29	4.47	5.20	7.00

As shown in Table 9, CAI from 2019 to 2023 shows different trends under five different growth rates of C-BI value. It can be seen more intuitively from Fig. 4. CAI increases when the C-BI value growth rate increases. CAI decreases when the C-BI value growth rate decreases. Moreover, CAI increases or decreases faster when the added value becomes larger or smaller. It indicates that the C-BI value has a significant effect on the value of CAI.

Fig. 4. CAI for different growth rates of C-BI

4.3 The relationship between R-BI and CAI

Table 10 shows the value of CAI and R-BI from 2013 to 2018. The prediction process is same as the Part 4.1. The results are shown in Table 11.

Table 10 the value of CAI and R-BI from 2013 to 2018

Year	2013	2014	2015	2016	2017	2018
CAI	5.26	7.56	6.78	6.08	5.57	5.04
new energy	158	223	287	299	331	383
surroundings protection	138	170	159	155	160	173
Environmental protection	212	242	277	247	290	266
Environmental Protection Agency	136	168	183	194	212	194
protect environment	107	152	154	186	200	226

air quality	668	1718	1936	1909	1955	2019
surroundings	130	173	172	172	178	189
Air purification	116	138	124	132	117	102

Table 11 CAI at different growth rates for R-BI value

Year	R-BI value increase rate				
	-10%	-5%	5%	10%	20%
2019	4.64	4.65	4.69	4.71	4.74
2020	4.25	4.31	4.44	4.51	4.66
2021	3.88	4.00	4.28	4.44	4.77
2022	3.52	3.72	4.19	4.46	5.07
2023	3.20	3.47	4.16	4.57	5.58

As shown in Table 11, CAI from 2019 to 2023 shows different trends under five different growth rates of R-BI value. It can be seen more intuitively from Fig. 5. CAI increases when R-BI value growth rate increases. CAI decreases when R-BI value growth rate decreases. Moreover, CAI increases or decreases faster when the added value becomes larger or smaller. It indicates that R-BI value has a significant effect on the value of CAI.

Fig. 5. CAI for different growth rates of R-BI

4.4 Comparative analysis

In the above process, the CAI simulation and prediction results were obtained through the Baidu index of hazards, causes and responses. The correlation degree between each keywords and CAI are at a high level. It shows that Baidu index can reflect CAI to a certain extent. The error is very small. So the grey model is very suitable for annual CAI prediction. Therefore, the prediction result has high reliability. According to the forecasting results, as the air quality-related Baidu Index grows faster, CAI becomes larger. CAI increase or decrease is great when the growth rates of the air quality-related Baidu Index is also at a higher or lower level. In other words, its speed of increase or decrease is increasing.

The overall trend is similar to H-BI. But compared to H-BI, C-BI greatly affects CAI. It can be seen

more intuitively from Fig.6 and Fig.7. They show that the influence of air pollution C-BI on CAI is dominant. The impact of R-BI on CAI is similar to that of H-BI. Based on this similarity, there may be some internal connections between them. According to the theory of protection motivation, perceived risk determines the willingness to respond. It means that the public also search for responses of air quality while search for hazards of air quality.

Fig. 6. The CAI value when the decrease rate of Baidu index is 10%

Fig. 7. The CAI value when the growth rate of Baidu index is 20%

5. Conclusion and implication

We use grey correlation analysis, grey multivariate convolution model and particle swarm optimization to discuss the relationship between CAI and Baidu index. According to the analysis results, as the air quality-related Baidu Index grows faster, air quality becomes worse. When Baidu index growth is at a high level, air quality deteriorates further. The relevant search for the cause of air quality can strongly reflect air quality. Searches for keywords such as smog, PM_{2.5}, PM₁₀, etc., all reflect the change of air

quality with a high degree.

Baidu index reflects the public's grasp of the air quality information and channels. Faced with air pollution, most people will take actions. They can be roughly divided into two categories: People who understand air pollution well can directly take corresponding actions and need not to search for related knowledge. People with few knowledge of air pollution will search for relevant knowledge. Obviously, the former can deal with air pollution better. These people are conducive to the improvement of air quality. Due to little or no understanding of air pollution, the response of the latter is reflected in searching air pollution-related information. When air pollution occurs, this group contributes little to improving air quality. So we can improve air quality by reducing the number of the latter population. Based on the above discussion, this article makes the following suggestions for improving air quality.

1. Improving the ability of the public perceive and respond to air pollution

Environmental risk perception can effectively help the public maintain a certain extent of sensitivity and understand the process, manifestations and coping methods of air pollution. When air pollution occurs, it can be detected immediately. The strengthening of the public's response knowledge not only helps to reduce self-exposure and avoid air pollution hazards, but also contributes to the improvement of air quality. The first barrier for prevent air pollution is formed through individual-centric prevention and control of all the people.

2. Strengthening effective communication among environmental protection departments and the public

Due to the complex information sources and different measurement methods, there is a certain discrepancy between the information released by all parties, which has caused public doubts. Air quality information is the main basis for the public to make decisions for air pollution. Therefore, the creation of

government's authoritative information release system and the construction of air pollution information communication channels should be completed as soon as possible. Only in this way can the public obtain air quality related information opportunistically and accurately. With relevant information about the nature, causes, and potential hazards of air pollution, the public can prevent and improve air pollution in a targeted manner.

3. Establishing public participation system for air pollution governance

The governance of air pollution requires the joint participation of government, non-governmental organizations, enterprises and the public. The public is both the demander and the promoter of the air governance process. Consequently, it is urgent to establish a subjective awareness of the public's active participation in air pollution governance. It contributes to motivate the public to take substantive action to protect the environment. For example, low-carbon travel, tree planting and energy conservation etc., which are all actions that help to improve air quality.

Acknowledgements

The relevant researches in this paper are supported by the National Natural Science Foundation of China (No.71871084), the Excellent Young Scientist Foundation of Hebei Education Department (No. SLRC2019001) and the project of high-level talent in Hebei province.

Funding

The relevant researches are supported by the National Natural Science Foundation of China (71871084) and the Excellent Young Scientist Foundation of Hebei Education Department (SLRC2019001).

Availability of Data and Material

All data in this paper is exposed and directly accessible.

Contributions

Leping Tu contributed to the conception and design of the study. Leping Tu conducted material preparation, data collection and analysis. The first draft of the manuscript was written by Yan Chen, and all authors reviewed early drafts of the manuscript. All authors read and approved the final manuscript.

Competing Interests

The authors declare that they have no competing interests.

Reference

- [1] Flatø H. Socioeconomic status, air pollution and desire for local environmental protection in China: insights from national survey data. *Journal of Environmental Planning and Management*, 2020; 63 (1).
- [2] Rebeca I, Saul GDS, Rafael B, la PDd, Denis S, Alberto G, Elena B. Health impact assessment by the implementation of Madrid City air-quality plan in 2020. *Environmental research*, 2020; 183 (
- [3] Paynesturges D, Marty MA, Perera F, Miller MD, Swanson M, Ellickson K, Coryslehta DA, Ritz B, Balmes JR, Anderko L. Healthy Air, healthy brains: advancing air pollution policy to protect children's health. *American Journal of Public Health*, 2019; 109 (4).550-554.
- [4] Xu H, Chen J, Zhao Q, Zhang Y, Wang T, Feng B, Wang Y, Liu S, Yi T, Liu S, Wu R, Zhang Q, Fang J, Song X, Rajagopalan S, Li J, Brook RD, Huang W. Ambient air pollution is associated with cardiac repolarization abnormalities in healthy adults. *Environmental Research*, 2019, 171 (
- [5] Wei-Ling W, Wen-Bo X, Yan-Li W, Yu L, Tao F, Ze-Lin C. Health benefit evaluation for air pollution prevention and control action plan in China. *Huan Jing Ke Xue*, 2019; 40(7).
- [6] Qi X, Mei G, Cuomo S, Liu C, Xu N, Data analysis and mining of the correlations between meteorological conditions and air quality: a case study in Beijing. *The internet of things*, 2019, pp. 100127.
- [7] Zeng Y, Cao Y, Qiao X, Seyler BC, Tang Y. Air pollution reduction in China: Recent success but great challenge for the future. *Science of The Total Environment*, 2019; 663:329-337.
- [8] Ahmad AK, Jafar A, Aljoumaa K. Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 2019; 6 (1).1-24.
- [9] Aqib M, Mehmood R, Alzahrani AI, Katib I, Albeshri A, Altowaijri SM. Smarter traffic prediction using big data, in-memory computing, deep learning and GPUs. *Sensors*, 2019; 19 (9).2206.
- [10] Lei MT, Monjardino J, Mendes L, Goncalves D, Ferreira FMFC. Macao air quality forecast using statistical

methods. *Air Quality, Atmosphere & Health*, 2019; 12 (9).1049-1057.

[11] Chen H, Lo T. Online search activities and investor attention on financial markets. *Asia-Pacific Management Review*, 2018; 24(1):21-26.

[12] Colladon AF, Guardabascio B, Innarella R, Using social network and semantic analysis to analyze online travel forums and forecast tourism demand. *Decision Support Systems*, 2019, pp. 113075.

[13] Lin H-Y, Yang S-Y. A cloud-based energy data mining information agent system based on big data analysis technology. *Microelectronics Reliability*, 2019; 97 (

[14] Lefebvre RC, Bornkessel AS. Digital Social Networks and Health. *Circulation* 2013; 127 (17).1829-1836.

[15] Zhang Y, Wang Y, Gao M, Ma Q, Zhao J, Zhang R, Wang Q, Huang L. A predictive data feature exploration-based air quality prediction approach. *IEEE Access*, 2019; 7:30732-30743.

[16] Zhao J, Deng F, Cai Y, Chen J. Long short-term memory-Fully connected (LSTM-FC) neural network for PM_{2.5} concentration prediction. *Chemosphere*, 2019; 220:486-492.

[17] Wen C, Liu S, Yao X, Peng L, Li X, Hu Y, Chi T. A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Science of the Total Environment*, 2019, 654:1091-1099.

[18] Li H, Wang J, Li R, Lu H. Novel analysis–forecast system based on multi-objective optimization for air quality index. *Journal of Cleaner Production*, 2019, 208:1365-1383.

[19] Chen Y, Wang J. Ecological security early-warning in central Yunnan Province, China, based on the grey model. *Ecological Indicators*, 2020; 111,106000.

[20] Zhang Y, Peng F, Mu J. Application of grey system theory on the corrosion behavior of steel in seawater. *Journal of The Institution of Engineers (India): Series C*, 2019, 100(4):693-699.

[21] Nguyen N-T, Tran T-T. Optimizing mathematical parameters of Grey system theory: an empirical forecasting case of Vietnamese tourism. *Neural Computing and Applications*, 2019, 31(2):1075-1089.

[22] Chen H-Y, Lee C-H. Electricity consumption prediction for buildings using multiple adaptive network-based fuzzy inference system models and grey relational analysis. *Energy Reports*, 2019, 5:1509-1524.

[23] Duman GM, Kongar E, Gupta SM. Estimation of electronic waste using optimized multivariate grey models. *Waste Management*, 2019; 95:241-249.

[24] Ma X, Xie M, Wu W, Zeng B, Wang Y, Wu X. The novel fractional discrete multivariate grey system model and its applications. *Applied Mathematical Modelling*, 2019; 70:402-424.

[25] Wu L, Gao X, Xiao Y, Yang Y, Chen X. Using a novel multi-variable grey model to forecast the electricity consumption of Shandong Province in China. *Energy*, 2018, 157:327-335.

[26] Ding S. A novel discrete grey multivariable model and its application in forecasting the output value of China's high-tech industries. *Computers & Industrial Engineering*, 2019, 127:749-760.

[27] Wang Z-X, Li Q. Modelling the nonlinear relationship between CO₂ emissions and economic growth using a PSO algorithm-based grey Verhulst model. *Journal of Cleaner Production*, 2019, 207:214-224.

[28] Zheng W, Li X, Xie J, Yin L, Wang Y. Impact of human activities on haze in Beijing based on grey relational analysis. *Rendiconti Lincei* 2015, 26(2).187-192.

- [29] Wu W, Ma X, Zhang Y, Li W, Wang Y. A novel conformable fractional non-homogeneous grey model for forecasting carbon dioxide emissions of BRICS countries. *Science of the Total Environment*, 2020,707,135447.
- [30] Wu L, Xu Z. Analyzing the air quality of Beijing, Tianjin, and Shijiazhuang using grey Verhulst model. *Air Quality, Atmosphere & Health*, 2019; 12(12).1419-1426.
- [31] Kothe EJ, Ling M, North M, Klas A, Mullan BA, Novoradovskaya L. Protection motivation theory and proenvironmental behaviour: a systematic mapping review. *Australian Journal of Psychology*, 2019,71(4): 411-432.

Figures

Figure 1

Location of Beijing. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

Figure 2

Flow chart of the study process

Figure 3

CAI for different growth rates of H-BI

Figure 4

CAI for different growth rates of C-BI

Figure 5

CAI for different growth rates of R-BI

Figure 6

The CAI value when the decrease rate of Baidu index is 10%

Figure 7

The CAI value when the growth rate of Baidu index is 20%