Research on China's Carbon Emission Prediction and Low-Carbon Development Strategy Based on Big Data

Huixin Man [#, *], Yanxiang Liu [#], Ying Cao [#]

School of Mathematics, Statistics and Mechanics, Beijing University of Technology, Beijing, China, 100124

* Corresponding Author Email: 21062319@emails.bjut.edu.cn

[#]These authors contributed equally.

Abstract. As the world's leading carbon emitter, China has a profound impact on global climate change. Based on data from 2000 to 2023, this paper constructs a comprehensive evaluation index system to evaluate the progress of China's "dual carbon" goal. Using entropy-weight TOPSIS method to determine the weight of each index, we found that the installed capacity of solar power generation and sales of new energy vehicles have the most significant impact on carbon emission. Since 2012, China's overall carbon emission score has continued to rise, highlighting the importance of emission reduction measures. In order to predict CO2 emission from 2024 to 2030, the GM(1,1) grey prediction model and BP neural network regression model are used in this paper. Although the relative error of the GM(1,1) model is only 1.769%, its single-variable prediction limits its persuasibility. Therefore, this paper adjusted the index system and built a BP neural network regression model, which successfully predicted the carbon dioxide emission from 2000 to 2030, with a model fitting degree of more than 0.9. The results show that the index evaluation system and prediction model can provide an effective assessment of China's carbon emission status and provide a scientific basis for the formulation of emission reduction policies.

Keywords: Big Data, Carbon emission prediction, entropy-weight TOPSIS method, BP Neural Network, Grey Forecast Model.

1. Introduction

1.1. Research background

With the increasingly serious problem of global warming, carbon emission has become the focus of global attention. In 2023, China accounted for the highest share of carbon dioxide emission in the world, reaching 37%, so China's carbon emission policies and mitigation measures have a significant impact on the global climate. August 15, 2023 is China's first Global Ecology Day, on this day, the National Development and Reform Commission said that China's carbon peak, carbon neutral "1+N" policy system has been completed, and continues to land. Since 2023, policies and systems related to "dual carbon" have been gradually improved, and the basic capacity of "dual carbon" work has been significantly enhanced [1]. China has committed to achieving carbon peak by 2030 and carbon neutrality by 2060. This is undoubtedly an important commitment to the global response to climate change [2-3]. Therefore, in order to achieve the goal of carbon neutrality and carbon peak and promote the building of a community with a shared future for mankind, it is necessary to scientifically evaluate the existing emission reduction policies and predict the change trend of China's future carbon emission and emission reduction.

1.2. Literature review

Global warming caused by greenhouse gas emission is an environmental issue of widespread concern in the international community. As a major carbon emitter, China is actively taking on emission reduction responsibilities commensurate with its international status, continuously increasing its Nationally Determined Contributions, committing to reaching peak carbon emission by 2030 and striving to achieve "carbon neutrality" by 2060 [4-5]. The implementation of low-carbon development strategies has irreplaceable significance for promoting production reform and economic

development. In this context, the problem of carbon emission forecasting continues to heat up. Traditional prediction models are mostly used for multivariate linear fitting. But this kind of model also has some problems, only has the ability of linear fitting, the fitting effect is average [6]. Therefore, in recent years, artificial intelligence algorithms have been introduced to make predictions, and the excellent nonlinear fitting ability of neural networks is very suitable for carbon emission prediction [7].

Existing literature has achieved rich research results in carbon emission forecasting. For example, many scholars have already conducted neural network carbon emission forecasting research in various regions and industries [7-10]. But many studies only use traditional or artificial intelligence algorithms, and relatively few combine the two. Therefore, this paper attempts to combine gray prediction and neural network algorithm to jointly predict the carbon emission results in the next 7 years, and can also give the predicted values of other relevant variables in the next 7 years.

1.3. Problem analysis

In order to better quantify the development of China's "double carbon" cause, this paper first established the "double carbon" index, using TOPSIS-entropy weight method to determine the weight of each index. The grey prediction model is established to make preliminary carbon emission prediction. Since the data are high-dimensional data, the grey prediction is used to screen variables, and the Bp neural network model is combined to predict the values of other variables, so as to make carbon emission prediction, explore the influence of other factors on carbon emission, and provide help and suggestions for the realization of the "double carbon" goal.

2. Construction and measurement analysis of "double carbon" index comprehensive evaluation system

2.1. Data source introduction

The "two-carbon" evaluation system is derived from the "Low-carbon Development Blue Book: China Carbon Neutral Development Report (2022)", most of which comes from the National Bureau of Statistics.

2.2. Construction of "double carbon" index comprehensive evaluation index system

This paper constructs a first-level index from two dimensions, which is represented by 29 second-level indexes under the first-level index, and constructs a complete evaluation system of "dual-carbon" indexes, as shown in Table.1.

Table.1. "Double carbon" index construction table

Primary index	Secondary index	Three-level index	Unit
Carburization index	Double control index	Total coal consumption	Ten thousand tons
		Energy intensity	Tons of standard coal / 10,000 yuan
		Carbon emission intensity	Tons of carbon dioxide / 10,000 yuan
	Power performance index	Share of thermal coal in coal use	%
		Installed thermal power capacity	megawatt
		Line loss rate	%
	Industrial carbon intensity index	Carbon emissions per unit of electricity	Tons of carbon dioxide / 10,000 KWH
		Carbon emissions per unit of steel	Tons of carbon dioxide/tons of crude steel
		Unit cement discharge	Tons of CO2 / cement production
		Unit coke emission	Tons of carbon dioxide/coke production
	Industrial high energy-consuming products capacity index	Crude steel output	Ten thousand tons
		Coke yield	Ten thousand tons
		Cement output	Ten thousand tons
		Flat glass output	Million TEU
Carbon reduction index	Green energy index	Installed capacity of wind power	Ten thousand kilowatt-hours
		Installed solar power capacity	Ten thousand kilowatt-hours
		Installed hydropower capacity	Ten thousand kilowatt-hours
		Installed nuclear power capacity	Ten thousand kilowatt-hours
	Green transport index	New energy tram sales	Ten thousand units
		Water passenger traffic	Thousands of people
	Circular economy index	Harmless disposal capacity of domestic waste	Ten thousand tons
	Ecological carbon sink index	Forest area	Million cubic meters
		Forest coverage rate	%
		Afforestation area	Ten thousand hectares
		Nature reserve area	Ten thousand hectares
		Soil erosion control area	Ten thousand hectares
	Low Carbon Agriculture Index	Energy consumption per unit of grain output	Tons of standard coal/tons of grain production
		Fertilizer use per unit of grain output	Tons of fertilizer used/tons of grain production

2.3. Index measurement based on entropy-weight TOPSIS method

2.3.1 Data preprocessing

The type of missing data in the dataset is random missing, and this paper use random forest method to supplement the missing values. Secondly, in order to unify the attributes of the indicators, this paper carries out positive data conversion and convert all negative indicators into positive indicators. Finally, this paper carries out max-min standardization of the data.

2.3.2 The entropy-weight TOPSIS method is used to calculate the weight

This paper calculates the weight of each indicator by entropy weight TOPSIS method. Among the 28 indicators, the installed capacity of solar power generation in terms of carbon reduction has the highest weight, reaching 12.825%, the sales volume of new energy trams is the second highest, with a weight of 11.417%, and the weight of the remaining indicators is lower than 5.3%, which is relatively not obvious.

The installed capacity of solar power generation, compared with other clean energy (wind energy, hydro energy, nuclear energy), solar energy resources are rich, especially in the northwest region of sufficient sunshine, strong ultraviolet light, can provide good natural conditions for power generation, and solar power technology progress and cost gradually reduce, while getting policy support, making solar power more competitive. Therefore, it has a very important role in carbon reduction.

Compared with traditional oil vehicles, new energy vehicles can reduce the emission of harmful gases such as carbon dioxide, have a higher energy conversion efficiency, and promote the use of renewable resources. From the perspective of policies and national conditions, China has adopted some policies to support the development of new energy vehicles, combined with China's large population and low cost of new energy vehicles, the sales of new energy vehicles have gradually increased, thus promoting the implementation of carbon reduction measures and achieving the reduction of carbon emission.

This paper draws the total weight table of primary and secondary indicators, as shown in Table.2.

Table.2. Total weight table of primary and secondary indicators

Primary index	Total weight of the Primary index	Three-level index	Total weight of Secondary index
Carbon reduction index	40.43%	Double control index	10.93%
		Power performance index	6.56%
		Industrial carbon intensity index	5.92%
		Industrial high energy-consuming products capacity index	17.02%
Carbon reduction index	59.57%	Green energy index	29.58%
		Green transport index	12.94%
		Circular economy index	3.86%
		Ecological carbon sink index	9.83%
		Low Carbon Agriculture index	3.36%

The weight of industrial energy-consuming products index is the largest in the carburization index. It can be seen that the emission of energy-consuming products such as crude steel, coke and cement have a greater role in promoting carburization. The weight of green energy index and green transportation index are respectively the first and second largest carbon reduction index, which shows that green energy such as solar power generation, wind power generation and green transportation modes such as new energy vehicles have a strong role in promoting carbon reduction.

The weight of carbon reduction index is higher than that of carbon increase index, indicating that carbon reduction index is more important in the model evaluation system and has a higher degree of differentiation among evaluation objects. It can be inferred that from 2000 to 2023, carbon reduction is the more critical influencing factor, so China's promotion of carbon reduction policies in recent years is effective.

2.3.3 Model result analysis

This paper visualized the index score, as shown in Figure 1.

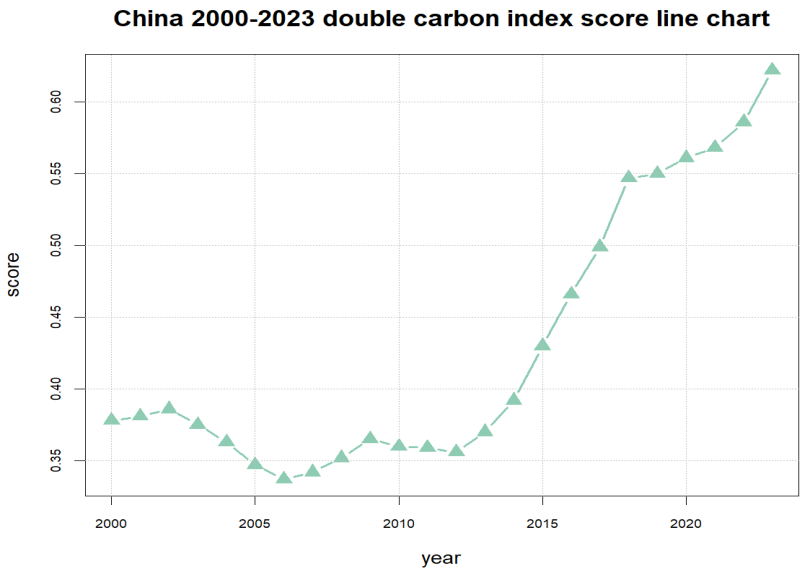


Figure 1. "Double Carbon" index score line chart 2000-2023

By analyzing the line chart, it can be seen that from 2000 to 2012, China's carbon emission index score showed a trend of fluctuating decline, and 2013 was the turning point of the score. From 2013 to 2018, China's carbon emission index score increased rapidly, and from 2018 to 2023, the growth rate of China's carbon emission index slowed down. But there is still an upward trend.

In 2013, China's carbon market was established, and the score index increased by 0.02; Since 2014, with the development of the market, the improvement of policies, and the government's promotion of energy conservation, emission reduction and new energy applications, carbon emission performance has improved year by year, and China's carbon emission performance has gradually risen from 2014 to 2023.

3. China's carbon emission forecast

3.1. Gray Forecast Model

3.1.1 Model building

Carbon emission in 2023 will decrease compared with 2022, so if data of earlier years is used, there will be great model error. Therefore, in order to reduce the relative error as much as possible, the data of 10 years from 2013 to 2023 will be selected only to forecast the data of the next seven years.

3.1.2 Model result

This paper uses 10 years of data from 2013 to 2023 to make forecasts for the next seven years. The fitting results are shown in Table.3.

Table.3. Grey Forecast Model fitting results table

Years	True Values/Mt	Predicted Values/Mt	Residuals/Mt	Relative Residuals/%
2013	9272.381	9272.381	0	0
2014	9284.947	8862.503	422.444	4.55
2015	9178.149	9104.471	73.678	0.803
2016	9171.746	9353.046	-181.3	1.977
2017	9376.098	9608.407	-232.309	2.478
2018	9831.926	9870.74	-38.814	0.395
2019	9992.189	10140.236	-148.047	1.482
2020	10087.694	10417.089	-329.395	3.265
2021	10682.83	10701.502	-18.672	0.175
2022	11477	10993.679	483.321	4.211
2023	11280	11293.834	-13.834	0.123

Below are the coefficient predictions made using MATLAB, as shown in Table.4:

Table.4. Coefficients in Gray Forecast Model

a	u
-0.0269	8493.9

Prediction model is as followed:

$$\hat{\mathbf{x}}^{(1)}(\mathbf{k}+\mathbf{1}) = \mathbf{3.25} \times \mathbf{10^5 e^{0.0269k}} - \mathbf{3.16} \times \mathbf{10^5} \quad (1)$$

The model test results are shown in the following Table.5.

Table.5. Accuracy test results

Relative residuals	Relative prediction error	Mean Square Error Ratio	Probability of Small Error
0.0177	0.0273	0.2994	1

Compare the predicted results with the original data, and visualize them as shown in Figure 2:

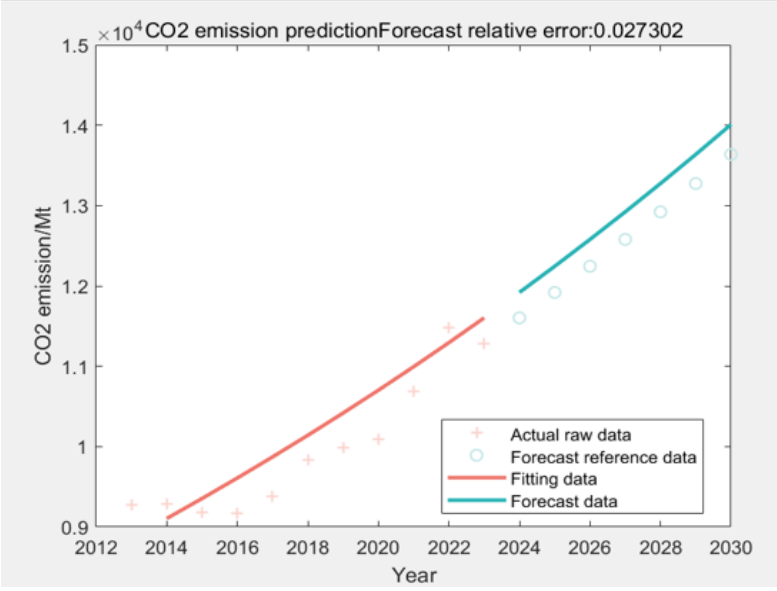


Figure 2. Gray Forecast Model prediction

Finally, this paper got the prediction, as shown in Table.6.:

Table.6. Gray Forecast Model predicting results table

Year	Predicted Values/Mt
2024	11602.18
2025	11918.95
2026	12244.37
2027	12578.67
2028	12922.10
2029	13274.90
2030	13637.34

3.1.3 Model conclusion

Based on the observation of the above chart and diagram, the carbon emission of this model will grow steadily in the next seven years and finally reach about 13.6 billion tons in 2030. Further predicting the carbon emission in 2031 according to the growth trend of this model may reach 14 billion tons, which is also consistent with the fact that the country leader announced at the United Nations General Assembly on September 22, 2020 that "China will reach its carbon peak by 2030 with carbon emission reaching 14 billion tons".

3.2. Bp neural network model

3.2.1 Modeling building

Because the grey prediction only predicts the value of one variable, carbon dioxide emission, but carbon dioxide emission is closely related to other tertiary indicators. Therefore, in order to improve the persuasive power of the model, this paper used the grey prediction method to screen out the other three indicators and predict their values in 2024-2030. Meanwhile, "carbon emission intensity" was replaced by "carbon dioxide emission". Finally, 24 variables were screened and predicted (the three indicators of total coal consumption, line loss rate and soil erosion control area were shaved).

In order to predict long-term carbon emission, this paper uses the error inverse propagation algorithm to train the neural network, and adjusts the connection weights by learning the patterns in the training data set.

3.2.2 Network structure and training process

Aiming at the dual-carbon index data from 2000 to 2023, 70% is taken as training set D.

$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}, x_i \in R^d, y_j \in R^l$, the input example has d attributes, the output 1-dimensional real-value vector, the hidden layer is the q layer. Where d=25, the number of hidden layers q is 10 according to Kolmogorov's theorem, combined with experience. Sigmod function is used for both the activation function of the hidden layer and the output layer. The form is as follows:

$$\mathrm{sigmod}(x) = \frac{1}{1 + e^x} \tag{2}$$

The structure of the neural network in the figure contains 271 parameters. The parameters are determined based on the gradient descent strategy, and the negative gradient direction of each unknown target is adjusted until the error reaches the set convergence value and the training stops.

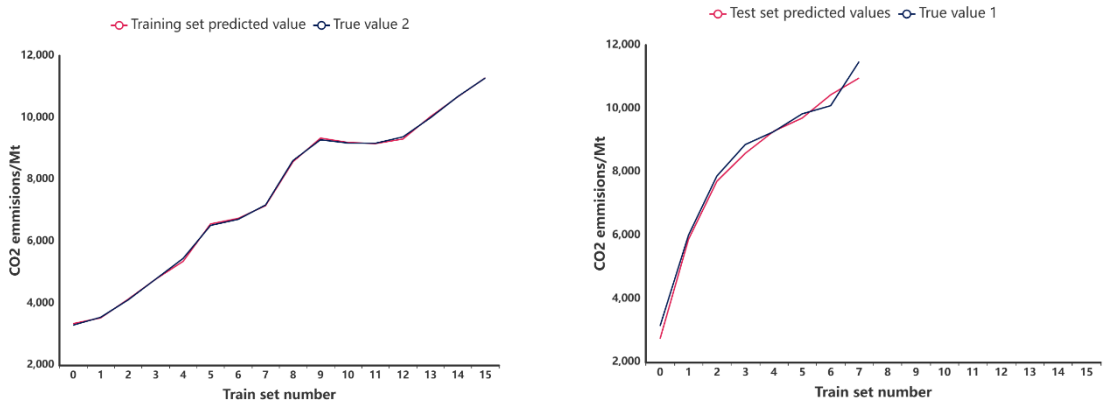
The error evaluation index is mean square error, and the learning rate η is 0.01, which is used to control the step size of iterative update. In addition, the cross-validation method is used to select the optimal model and reasonable super parameters. In order to prevent overfitting, a penalty term is added, L2 regularization is performed on the weight vector, and a regular term is added to the loss function to control the training process.

3.2.3 Model result

The predicted and true values were plotted on the training set and the prediction set respectively, and the fitting prediction effect was compared. According to the analysis of Figure 3, it can be seen

that the prediction effect of the model is very good in both the test set and the training set, and it can predict the carbon emission more accurately.

The forecast results for 2024-2030 are shown in Figure 3. (a) and Figure 3. (b):



(a) Train set prediction (b) Test set prediction

Figure 3. Train set prediction and Test set prediction

From the above forecast results, it can be seen that the growth rate of China's carbon emission will gradually slow down. The forecast results of carbon emission in 2024-2030 are shown in the table below. It can be seen that the carbon emission will increase year by year in the next six years and finally reach the target of 14073.08Mt in the year of "carbon peak". The results of neural network prediction model are shown in Table.7.

Table.7. Prediction results of neural network model

Year	2024	2025	2026	2027	2028	2029	2030
Forecast result /Mt	11661.26	12022.18	12398.39	12790.71	13199.99	13627.13	14073.08

The following table shows the evaluation indicators of relevant models. In order to exclude the influence of other dimensions, the data are normalized. The model test is shown in Table.8.

Table.8. Model check table

	MSE	RMSE	MAE	MAPE	R^2
Training set	0.0003	0.055	0.0046	2.8%	0.999
Test set	0.01	0.1	0.038	5.8%	0.986

It can be seen from the table that MSE is lower than 0.1, indicating high prediction accuracy. In addition, RMSE, MAE and MAPE all performed well in the test set, and the model generalization ability was strong. The R2 of the test set reached 0.986, and the model fitting effect was good. In addition, the MAPE values of the training set and the test set are both lower than 10%, which indicates that the overall performance of the model is good.

All the predicted variables in the next seven years are divided into four categories according to the size of the order of magnitude, and the area map is drawn respectively. Combined with the calculated annual growth rate, the changes of different variables in different years are analyzed. It can be seen that most of the variables show an overall growth trend, and some of the carbon increase index variables are as follows: Cement production, coke production, etc., its future increment will slow down, the increase rate is below 1%, and the energy intensity will show negative growth in the future and continue to accelerate the rate. In addition, some carbon reduction indexes, such as new energy vehicle sales (growth rate of 6.75%), solar power generation installed capacity (10.27%), wind power installed capacity (10.27%), etc., have a large growth rate, and the growth rate is relatively stable.

4. Research conclusions and policy recommendations

4.1. Research conclusions

First, this paper studied each indicator based on the entropy weight TOPSIS method and obtained the weight values of each indicator. Among them, the installed capacity of solar power generation and the sales volume of new energy electric vehicles had the highest weight percentages, both exceeding 10% of the total. Secondly, this paper scored the carbon emission performance over the years, and the overall score has shown an upward trend since 2012. Finally, in view of the upcoming peak carbon emission in 2030, this paper used the grey prediction model and BP neural network model to predict carbon emission respectively. The two models had good test results, indicating their excellent predictive and practical capabilities. Moreover, the prediction results of the two models were 13.637 billion tons and 14.073 billion tons, respectively, which were very close to the level of 14 billion tons announced at the United Nations General Assembly, indicating that the establishment of the models was very successful.

4.2. Policy recommendations

Suggestions for *carbon* reduction targets

The government has strengthened the promotion of new energy vehicles and increased public awareness through incentives such as car purchase subsidies and tax incentives recently. But at the same time, this paper also encourages the government to invest for the solar energy industry, providing financial subsidies and tax incentives and lowering the investment threshold. Promoting the collaboration between industries and encouraging enterprises to innovate in order to achieve clean energy self-sufficiency are both of great importance.

(1) Suggestions for carbon-increasing indicators

This paper suggests the revision of the Coal Law and appeals to the coal chemical industry for aiming at the high-end, diversified and low-carbon development in order to advocate the clean and efficient use of coal. Moreover, the government should focus on the coupling development of fossil fuel and new energy, and give policy to reduce energy consumption and optimize energy structure.

(2) Suggestions based on the comprehensive scores of previous years and the trend before carbon peaking

Based on the carbon emission of the previous year, it is suggested to maintain stability, proceed in an orderly way and respond flexibly every year.

References

[1] Bai Xue. The progress of China's green transformation from the "double carbon" work in 2023 [J]. China Development and Reform Newspaper, 2024(1):1.

[2] Duan Hongbo, Wang Shouyang. China's challenge: Strategic adjustment of global temperature control target from 2 ° C to 1.5 ° C [J]. Management World, 2019, 35(10):50:63.

[3] Cai Bofeng, Cao Libin, Lei Yu, et al. Carbon dioxide emission path under China's carbon neutrality goal [J]. China Population, Resources and Environment, 2019,31(1):7-14. (in Chinese)

[4] Duan H B, Wang S Y. China's challenge: Strategic adjustment of global temperature control targets from2 ℃ to 1.5 ℃ [J]. Managing the World, 2019, 35(10):50:63.

[5] Romanello M, Whitmee S, Mulcahy E, et al. Further delays in tackling greenhouse gas emission at COP28 will be an act of negligence[J]. The Lancet, 2023, 402(10417): 2055-2057.

[6] Zhan J G, Shao Z. Research on Carbon Emission Forecasting Model for Cement Enterprises Based on Data-Driven[J]. Journal of Wuhan University of Technology, 2024, 46(03):32-39.

[7] Acheampong A O, Boateng E B. Modelling carbon emission intensity: Application of artificial neural network[J]. Journal of Cleaner Production, 2019, 225: 833-856.

[8] Deng W Z, Xu Z J. Characteristics of agricultural carbon emission and carbon peak analysis in Hunan Province[J]. Chinese Journal of Eco-Agriculture, 2024, 32(2): 206−217.

[9] Yang J Q, Fan X J, Zhao Y H, et al. Prediction of carbon emission in Shanxi Province based on PSO-BP neural network[J]. Journal of Environmental Engineering Technology, 2023,13(6):2016-2024.

[10] Hu Z R, Zhao W F, Song Y X, et al. Prediction spatial distribution of soil organic matter based on improved BP neural network with optimized sparrow search algorithm[J]. Environmental Science, 2024,45(5).