Prediction Of Global Temperature Based on ARIMA And GM Model

Jingyan Ye [1], Xuezhe Gao [1], Liang Wu [1], Ailin Zhang [2], Peng Wu [3], Xiang Yu[4, *]

[1] School of Materials Science and Technology China University of Geosciences (Beijing) China, 100083

[2] School of Economics and Management, China University of Geosciences (Beijing) China, 100083

[3] School of Materials Science and Technology China University of Geosciences (Beijing) China, 100083

[4] China University of Geosciences (Beijing) China, 100083

* Corresponding Author Email: yuxiang@cugb.edu.cn, 1003200601@cugb.edu.cn, 1003200607@cugb.edu.cn, 1003194123@cugb.edu.cn, 1007200102@cugb.edu.cn, 1003200404@cugb.edu.cn

Abstract. Global temperatures have risen rapidly in recent decades .In order to find the rising trend of global temperature, this paper establishes ARIMA model and GM model to fit the past temperature changes and predict the future temperature trend. The results showed that the global average temperature will reach 23.55 °C by 2050. Then through the evaluation index SSE, it is judged that the ARIMA (0,1,2) model hasthe best prediction effect.

Keywords: ARIMA, GM model, global temperature forecast.

1. Introduction

Global warming is the rise in global temperature.We learned that greenhouse gases such as carbon dioxide have high visible light transmittance of solar radiation and high absorption of long-wave radiation reflected from the earth,which lead to an increase in earth energy. We call this phenomenon the greenhouse effect. On the one hand， human factors have increased carbon dioxide emissions and causing climate warming. On the other hand, natural disasterssuch as marine geological changes, earthquakes, volcanic eruptions, forest fires, and virusesalso have a certain impact on global climate changes [1].

As is known to all, global warming not only makes many countries set a temperature record and affect the normal life of the citizens, but also endangers the natural ecosystem andthreatens the survival of human beings [2]. Therefore, Understanding the law of global climate change and its influencing factors is significant for us to carry out scientific and reasonable countermeasures.

By reading a lot of relevant literature, we found that predecessors have used LSTM long short-term memory network [3], BP neural network [4] and GA-BP neural network [5] to establish mathematical models and predict temperature. In this paper, we use ARIMA and GM model, and use Python, MATLB, SPSS software to program the above mathematical models.

We did the following research on global temperature change. First, we preprocessed the data we collected and plotted a up and downsline chart. Secondly, based on the datasets we processed, we built two mathematical models to describe the past and predict the future global temperature level. Finally, we used each of our models to predict global temperatures in 2050 and 2100 respectively and found out the best model.

A graphic of our work is shown by Figure 1:

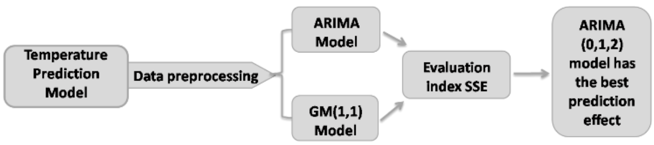

Fig. 1 Graphic of our work

2. Temperature Prediction Model

In this section, we preprocess the attached data first, then based on the processed data, we use two models to describe the temperature level and predict the average global temperature.

2.1. Data Preprocessing

Before data analysis, the availability of data must be guaranteed.

• *Missing value processing.*

From the Berkeley Earth Data page, we can easily get the regional temperature values from 1833 to 2022.At the same time, we found that temperature values were missing for some months and we considered rounding and completing them.We found that there is a large amount of missing data for the years before 1899. From the perspective of social development history, there was no significant global warming trend because it was before the second industrial revolution and there was no significant use of energy sources such as gas and oil. And the temperature measurement tools at that time was not advanced, there were many errors. Therefore, we remove all data before 1899.

•*Data average calculation.*

For the data of different sampling points obtained above, the method of directly averaging the temperature of all sampling points is not applicable because of the difference in their latitude and longitude, which will erase the key information of the location of the sampling points. Therefore, we divide the data into three types: southern temperate zone, northern temperate zone, and tropical zone, and average the data of all sampling points in the southern temperate zone, northern temperate zone, and tropical zone in each month to represent thetemperature of the corresponding temperature zone in that month. Under the above operation,we calculated the global average temperature and its growth rate from 1990 to 2022.

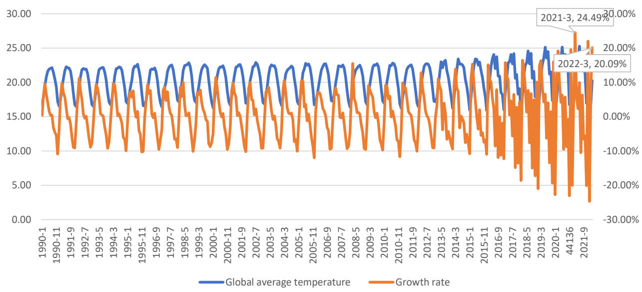

Fig. 2 Ups and downsline chart

As shown in the Figure 2, the blue curve is the monthly average global temperaturefrom January 1990 to March 2022, and the orange curve is the growth rate of this month relative to the previous month during January 1990 to March 2022. We can seethat the growth rate in March 2021 is higher than March 2022.

2.2. The Establishment of ARIMA Model

Improving forecasting especially time series forecasting accuracy is an important yet often difficult task facing decision makers in many areas. Autoregressive integrated moving average (ARIMA) models are one of the most popular linear models in time series forecasting,which have been widely applied in order to construct more accurate hybrid models during the past decade.[6].

• Autoregressive (AR). Describe the relationship between the current value and the historical value and use the historical time data of the variable to predict itself. The order of ARmodel is recorded as p value.[7] The formula is as follows:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \varepsilon \tag{1}$$

Where μ is the constant, p is the order, γ_i is the autocorrelation coefficient, ε is the error, y_t denotes the dependent variable for the forward t steps.

• Integrated(I). When the time series becomes stationary, the difference needs to be made, and the order of the difference is recorded as d value. Generally, first order is enough.

• Moving average (MA). Moving average model focuses on the accumulation of error terms in autoregressive model. The order of MA model is recorded as q value.[8] The formula is as follows:

$$y_t = \mu + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \varepsilon_t \tag{2}$$

Where μ is a constant, q is the order, β_i is the delay coefficient, ε_t is the white noise sequence, y_t denotes the dependent variable for the forward t steps.

The expression of the ARMA autoregressive moving average model is as follows:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \varepsilon_t \tag{3}$$

Letters mean the same as above.

2.3. The Establishment of GM（1，1）Model

Gray prediction, which refers to the prediction of the development of changes in the values of the characteristics of the system behavior, is carried out for systems containing both knownand uncertain information, that is, for gray processes that vary within a certain range and arerelated to time series. Although the phenomena shown in the gray process are random andhaphazard, they are, after all, ordered and bounded, so the obtained data set possesses potentiallaws. Gray prediction is the use of such laws to build gray models for predicting gray systems.[9]

The most widely used gray prediction model is the one-variable, first-order differential GM (1,1) model on series prediction. It is based on a random original time series, and the new timeseries formed by accumulating by time presents a pattern that can be approximated by the solution of the first-order linear differential equation. It is proved that the original time seriesrevealed by the approximation of the solution of the first-order linear differential equation hasan exponential variation pattern.[10] Therefore, the prediction of the gray model GM (1,1) is verysuccessful when the original time series implies an exponential variation law.

The procedure for solving the predicted values is as follows:

Let $X^{(0)}$be the GM (1,1) modeling sequence:

$$X^{(0)} = (x^0(1), x^0(2), \dots, x^0(n)), \tag{4}$$

$X^{(0)}$ is a 1-AGO (once-accumulated generation) sequence of $X^{(1)}$:

$$X^{(1)} = (x^1(1), x^1(2), \dots, x^1(n)), \tag{5}$$

$$x^{(1)}(k) = \textstyle\sum_{i=1}^{k} x^{(0)}(i), k = 1,2\dots, n \tag{6}$$

Let $Z^{(1)}$ be the average of the immediate neighbors (MEAZ) of $X^{(1)}$to generate the sequence.

$$Z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)) \tag{7}$$

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1) \tag{8}$$

Then the grey differential equation of GM (1,1) is as follows:

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{9}$$

Where a is called the development coefficient and b is the gray action quantity. Let $\widehat{\alpha}$ bethe parameter vector to be estimated, namely $\hat{a}=(a,b)^T$, then the least-quares estimated parameter column of the gray differential equation satisfies the following equation:

$$x^{(0)}(k)+az^{(1)}(k)=b \tag{10}$$

In addition,

$$B=\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \cdots & \cdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, Y_n \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \cdots \\ x^{(0)}(n) \end{bmatrix} \tag{11}$$

We call.

$$\frac{dx^{(1)}}{dt}+ax^{(1)}=b \tag{12}$$

the whitening equation of the gray differential equation

$$x^{(0)}(k)+az^{(1)}(k)=b \tag{13}$$

The predicted value can be obtained as above:

$$\hat{x}^{(1)}(k+1)=\left[x^{(1)}(0)-\frac{b}{a}\right]e^{-ak}+\frac{b}{a}, k=1,2,\cdots,n \tag{14}$$

Finally, we performed a $\hat{x}^0(i)$ residualsize test, i.e., a point-by-point test on the residuals of the model and actual values. First, $x^{(0)}(i)$ $\hat{x}^{(1)}(k+1)$ is calculated by the modelvalue , and $\hat{x}^0(i)$ $\hat{x}^{(1)}(k+1)$ is generated by cumulative subtraction. Finally, the absolute and relative residuals of the original series and are calculated, and the average relativeresiduals are calculated.

Absolute residual series:

$$\Delta^{(0)}=\{\Delta^{(0)}(i), i=1,2,\dots,n\}, \Delta^{(0)}=|x^{(0)}(i)-\hat{x}^{(0)}(i)| \tag{15}$$

Relative residual series:

$$\varphi=\{\varphi_i, i=1,2,\dots,n\}, \varphi_i=\left[\frac{\Delta^{(0)}(i)}{x^{(0)}(i)}\right]\% \tag{16}$$

Average relative residuals:

$$\bar{\varphi}=\frac{1}{n}\sum_{i=1}^{n}\varphi_i \tag{17}$$

2.4. The Solution of Models

2.4.1. Describe Past Global Temperature Levels

The descriptive statistical chart Table Ishows that the minimum value of temperature during the 122 years 1990-2021 is 18.83 degrees Celsius, the maximum value is 21.14 degrees Celsius, and the mean value is 19.74 degrees Celsius, with a variance of 0.228, indicating that the annual average global temperature is in a relatively stable state.

Tab. 1 Descriptive Statistics

Descriptive Statistics						
	N	Minimum	Maximum	Mean	Std.Deviation	Variance
Average T	122	18.8389668	21.1444024	19.7387573	0.47791289	0.228
ValiN (listwise)	122					

From the following plot of global annual mean temperature versus time Figure 3, it can be seenthat there is an increasing trend in global annual mean temperature, and the increasing trend ismore obvious in recent years.

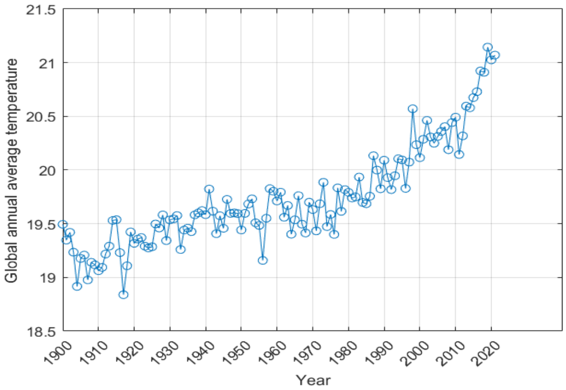

Fig. 3 Predictions of global temperatures

2.4.2. ARIMA model prediction results

We first selected the ARIMA (0,1,2) time series model based on the BIC criterion.

The ACF and PACF graphs of the residuals in Figure 4 show that the autocorrelation and partial autocorrelation coefficients for all lag orders are not significantly different from 0.

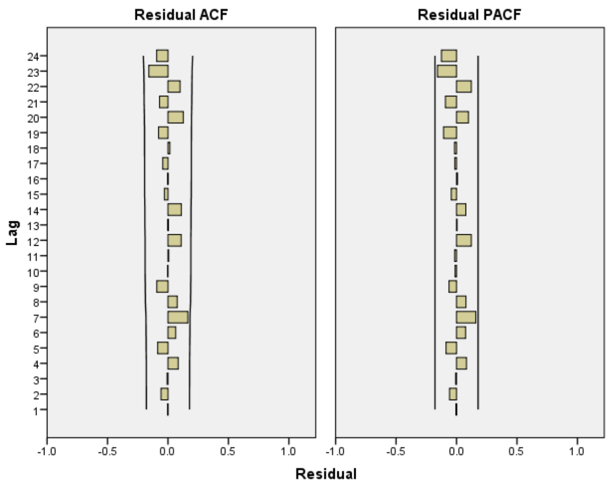

Fig. 4 Residual

Also, from the table II, we can see that the p-value obtained from the Q-test on the residuals is 0.746, i.e., we cannot reject the originalhypothesis that the residuals are white noise series, so the ARIMA (0,1,2) model is able to identify the temperature data in this case very well.

Tab. 2 Model statistics

Model Statistics						
Model	Number of Predictors	ModelFit statistics	Ljung-Box Q (18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Average T	0	0.367	11.973	16	0.746	2

In the following tableIII we can derive the parameters of ARIMA (0,1,2)

Tab. 3 ARIMA Model parameters

ARIMA Model Parameters					
		Estimate	SE	t	Sig.
Constant		0.009	0.003	3.318	0.001
Difference		1			
MA	Lag1	0.59	0.088	6.672	0
	Lag2	0.229	0.088	2.604	0.01

The model parameters are then brought into the following equation:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \varepsilon_t \tag{18}$$

where, μ is a constant 0.009; p is 0; q is 2; β_1 is 0.590;β_2 is 0.229; and ε_t is the perturbation term.

The fitted curve of the ARIMA (0,1,2) model and the predicted curve for the future aregiven in the following figure 5:

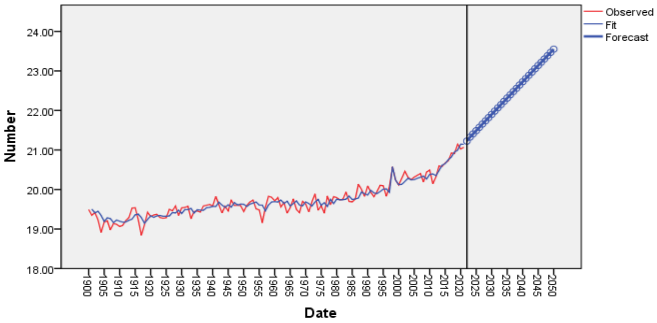

Fig. 5 The fitted and predicted curve

2.4.3. GM(1,1) Model Prediction Results

GM (1,1) model of Quasi-exponential law test:

Indicator 1: The proportion of data with smooth ratio less than 0.5 is 99.1736%>60%.

Indicator 2: Excluding the first two periods, the proportion of data with smoothness ratioless than 0.5 is 100% > 90%, so the data in this case can pass the test.

The figure 6 shows the smoothness of the original data, we can see that almost all points are. below the threshold, i.e., the smoothness ratio is almost all less than 0.5.

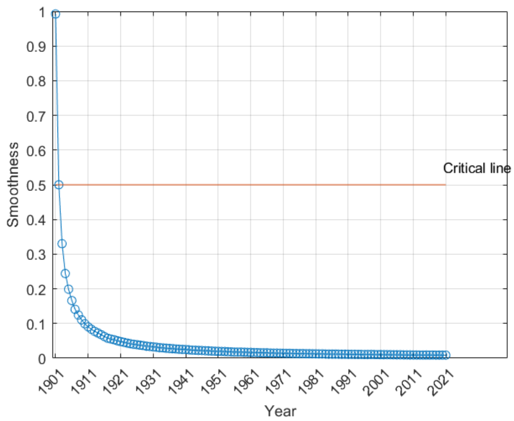

Fig. 6 Smoothness of the original data

Since there are three types of GM (1,1) models, the model with the smallest SSE can beselected as the optimal model by dividing the original data into experimental and test groups (3 groups of data, respectively, 2019, 2020 and 2021 data) and using SSE as the evaluation index.

The error sum of squares of traditional GM (1,1) for the test group prediction is 1.425.

New information GM (1,1) for the test group predicts the error sum of squares of 1.425.

The sum of squares of the errors predicted by the metabolicGM(1,1) for the test group was1.4034

Because the metabolic GM (1,1) model has the smallest error sum of squares, we shouldchoose it for prediction. This result is represented in the figure 7

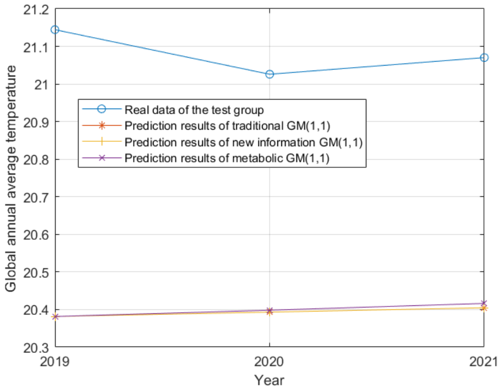

Fig. 7 Prediction results

The relative residuals, and grade deviation plots are shown in Figure 8. The evaluationresults of the fit of the metabolic GM (1,1) model to the original data will be output below.

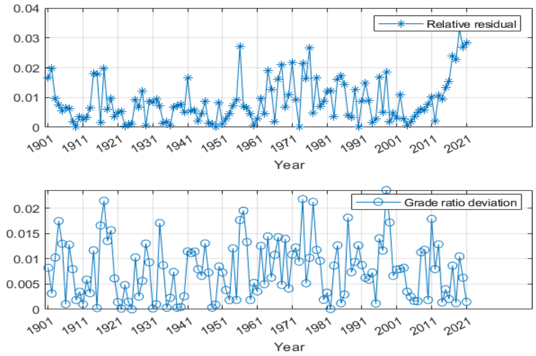

Fig. 8 The relative residualsand grade ratio deviation

The average relative residual was 0. 0087533.The results of the residual test show that themodel fits the original data very well

The average stepwise deviation is 0. 0077757.The results of the cascade bias test show thatthe model fits the original data very well

The fitted curve of the metabolic GM (1,1) model and the predicted curve for the future aregiven in the figure 9.

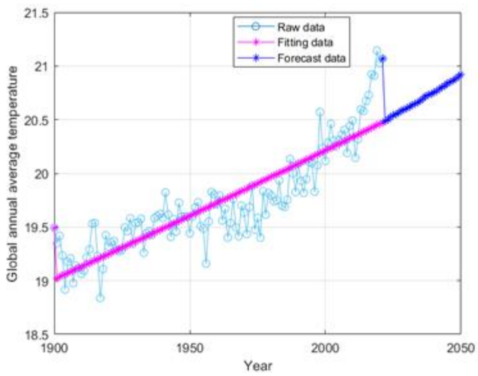

Fig. 9 The fitted curve and the predicted curve

Finally, the results of our future predictions using ARIMA (0,1,2) model, and metabolicGM(1,1), respectively, are shown in the tableIV:

Tab. 4 Prediction results

	ARIMA(0,1,2)	Metabolic GM(1,1)	New message GM(1,1)	Traditional GM(1,1)
SSE	0.0431	1.4034	1.425	1.425

Tab. 5 SSE result

Year	ARIMA	GM(1,1)		Year	ARIMA	GM(1,1)
2023	21.32	20.5		2037	22.48	20.72
2024	21.41	20.52		2038	22.56	20.73
2025	21.49	20.54		2039	22.64	20.74
2026	21.57	20.55		2040	22.73	20.75
2027	21.65	20.57		2041	22.81	20.77
2028	21.74	20.58		2042	22.89	20.79
2029	21.82	20.59		2043	22.97	20.8
2030	21.9	20.61		2044	23.06	20.82
2031	21.98	20.62		2045	23.14	20.83
2032	22.07	20.63		2046	23.22	20.85
2033	22.15	20.65		2047	23.3	20.86
2034	22.23	20.66		2048	23.39	20.88
2035	22.31	20.68		2049	23.47	20.9
2036	22.4	20.7		2050	23.55	20.92

According to the table IV, it can be seen that the

ARIMA (0,1,2) model, and the metabolic GM (1,1) model are consistent with the prediction that the global average temperature will reach 20 degrees Celsius in 2050 at the observation point.

Also, we believe that the ARIMA (0,1,2) model is the most accurately constructed. We divided the original data into an experimental group and a test group, and the predicted values of the test group were compared with the actual values, and the SSE of each model was obtained as table V: we can see that the ARIMA (0,1,2) model has the smallest SSE, so we think that the ARIMA (0,1,2) model is the most accurately constructed.

3. Strengths and Weaknesses of Our Models

3.1. Strengths

The ARIMA model used in our prediction ensures the stability of time series data.

We use three GM (1,1) models and selects the metabolic GM (1,1) model with the best prediction effect.

3.2. Weaknesses

The G M (1,1) model cannot explain the uncertain relationship between the factors in the system.

3.3. Further Discussion

Our improvements to the model are as follows:

The GM-ARIMA model is used to solve the disadvantage that the GM (1,1) model cannot explain the uncertain relationship between the factors in the system.

4. Conclusion.

Through the ups and downs line chart, we found that March 2022 is not the fastest growing month of temperature. After that, we use the ARIMA (0,1,2) model and the metabolic GM (1,1) model to predict the future global temperature, and predict that the global average temperature will reach 23.55 °C by 2050, which is consistent with the prediction that the global average temperature will reach 20 °C by 2050. Then through the evaluation index SSE, it is judged that the ARIMA (0,1,2) model has the best prediction effect..

References

[1] Global temperature rise may be higher than previously predicted [J]. Heilongjiang Science and Technology Information, 2012 (11) : 12-13.

[2] QIAN Weihong. Regularity, cause and future forecast of global temperature change [C] / /. Proceedings of the Subtropical Monsoon and Climate Change Branch of the 27th Annual Meeting of China Meteorological Society. [Publisher unknown], 2010: 80-81.

[3] Wang Yuanhao. Global temperature prediction analysis based on ARIMA model and LSTM neural network [J]. Scientific and technological innovation, 2021 (35): 166-170.

[4] Hou Huiqing. Global climate change prediction model based on BP neural network [J]. Technology and innovation, 2021 (09): 10-11.DOI : 10.15913 / j.cnki.kjycx.2021.09.005.

[5] Yuan Ming, Shen Jinshan, Zhang Yi, Yang Ruijia, Zhang Yunhui. Global temperature prediction based on GA-BP neural network [J]. Journal of Zhengzhou Aeronautical Industry Management College, 2020,38 (06) : 67-75.DOI : 10.19327 / j.cnki.zuaxb.1007-9734.2020.06.007.

[6] Ma Xia, Zhang Jinzhu, Li Can, Guo Zunguang. Application of ARIMA model teaching case design of R software [J]. Fujian Computer, 2022,38 (11) : 113-117.DOI : 10.16707 / j.cnki.fjpc.2022.11.026.

[7] Li Bingxiao, Zhang Shiwei, Zheng Shuyu, Zhao Zhifan. Research on hydropower power prediction based on multiple linear regression and ARIMA combination model [J]. Scientific and technological innovation, 2022 (33) : 71-74.

[8] Construction and numerical simulation of SVM-ARIMA model for CPI prediction [J / OL]. Statistics and decision-making, 2022 (21) : 48-52 [2022-11-27].DOI : 10.13546 / j.cnki.tjyjc.2022.21.009.

[9] Gu Yanchun, Lu Haiyan, Zhao Jinjin, Xu Jie, Shen Guanqiang. Improved GM (1,1) model based on chicken swarm optimization algorithm [J]. Information technology and informatization, 2022 (10) : 216-220.

[10] Guo Yang. Prediction of Main Indicators of River Water Quality Based on GM (1,1) Model [J]. Agriculture and Technology, 2022, 42 (19) : 116-120.DOI : 10.19754 / j.nyyjs.20221015029.