Global Warming Research Based on GA-BP Neural Network Model

Yichen Li [#,*] , Ce Lyu [#], Wei Xu [#]

School of Computer Science and Technology, Shandong University, Qingdao, China, 266237

* Corresponding Author Email: 202000130119@mail.sdu.edu.cn

[#]These authors contributed equally.

Abstract. Since the industrial revolution, global carbon emissions have continued to increase, and after entering the 21st century, the trend of global warming has become more and more obvious. This paper first use the collected 10-year monthly average temperature data and divide them into two groups: 2012.3~2022.2 and 2022.3~2022.10, then conduct the Mann-Whitney U test, and find that the temperature increase in March 2022 did not increase significantly; then we three models are established to predict the future global temperature, the first model is ARIMA, the second model is BP neural network, and the third model is GA-BP neural network. Three models are used to predict the time when the future temperature will reach 20°C, ARIMA predicts the time is 2050, BP is 2046, and GA-BP is 2045. Finally, this paper use MAE, MAPE, and RMSE to evaluate the accuracy of the model, and find that GA-BP has the highest accuracy, followed by BP, and ARIMA is the worst.

Keywords: ARIMA, BP neural network, GA-BP neural network, ridge regression, partial least squares regression.

1. Introduction

Global warming is a phenomenon related to natural environmental factors [1]. The cause of global warming is the continuous accumulation of the greenhouse effect, which leads to an imbalance in the energy absorbed and emitted by the earth-atmosphere system, and the energy continues to accumulate in the earth-atmosphere system, which leads to a continuous rise in the temperature of the earth's surface and global warming [2-4]. In recent years, the trend of global warming has intensified. Especially this year, the temperature in many places around the world has hit record highs. The Valley of Death in California, USA set a record high of 54.4°C in the history of human beings to record the earth's temperature [5]. Before the industrial revolution, the concentration of carbon dioxide in the atmosphere was around 280ppm. After entering the 21st century, the concentration of carbon dioxide in the atmosphere continued to rise. In 2004, the concentration of carbon dioxide in the atmosphere once exceeded the record of the largest 10-year average increase [6, 7].

The objective of this study is (A) Based on the attached data and related historical data sets collected, establish at least two mathematical models to fit past global temperature changes and predict future global temperature change trends. (B) Use one of the models given in (b) to predict the global average temperature in 2050 and 2100, respectively. Use the prediction results of the model to analyze whether the established model agrees that the global average temperature will reach 20°C in 2050 or 2100. If it is not the case that the model disagrees, give the time frame when the model predicts that the global average temperature will reach 20 degrees Celsius. (C) Analyze the accuracy of the model in (b) and give a reasonable explanation.

2. Data preprocessing

Due to the large amount of data, when trying to collect data related to global temperature, it is difficult to accurately find the accurate data of the relevant indicators of the corresponding date of global temperature, and the data given in the title also has a large number of missing data. In order to build a more accurate model, we interpolate the missing data using cubic spline interpolation algorithm.

3. Results and analysis

3.1. ARIMA model building

Problem A asks us to build predictive models to describe the past and predict future global temperatures. The data we collected has obvious time series characteristics, so we firstly use the ARIMA model (Figure 1).

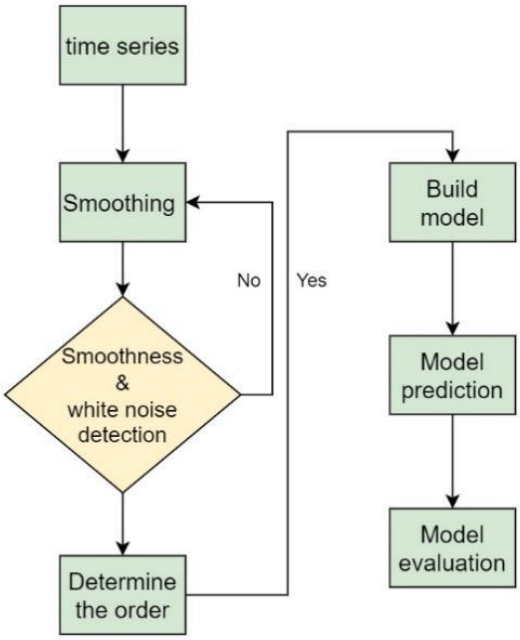


Figure 1. ARIMA modeling process

3.1.1 ARIMA model preparation

Differential Autoregressive Moving Average Model (ARIMA).

The ARIMA (p, d, q) model converts the non-stationary time series into a stationary time series, and then the dependent variable only regresses its lag value and the current value and lag value of the random error item; the formula is as follows:

$$y_t' = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-1}' + \varepsilon_t \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} + \varepsilon_t \quad (1)$$

y_t is the current predicted value, α_0 is the constant term, α_i is the autocorrelation coefficient ε_t is the white noise sequence (error), p is the autoregressive item; q is the moving average item, and d is the number of differences made when the time series becomes stable.

3.1.2 Solution of ARIMA model

Step1: Smoothing

Observe the time series map of the global land annual average temperature from 1959 to 2022. The global continental temperature has an upward trend, and the data stability is poor. Therefore, we should perform differential processing on the original data (Table 1).

Table 1. ADF checklist

variable	difference order	t	P	AIC	critical value		
					1%	5%	10%
temperature	0	0.715	0.990	-24.931	-3.571	-2.923	-2.599
	1	-6.426	0.000***	-26.427	-3.571	-2.923	-2.599

Note: ***, **, * represent the significance levels of 1%, 5%, and 10% respectively.

Through the ADF test, it is known that when the difference is 1st order, the significant P value is 0.000, the level is significant, and the null hypothesis is rejected. The sequence is a stable time series, which is consistent with the intuitive feeling. That is to say, the stability of the original data is enhanced after the first-order difference, and the value of the parameter d is 1 model.

Step2: Checking

SPSS automatically finds the optimal parameters based on the AIC information criterion, and the model result is an ARIMA model (0, 1, 1).

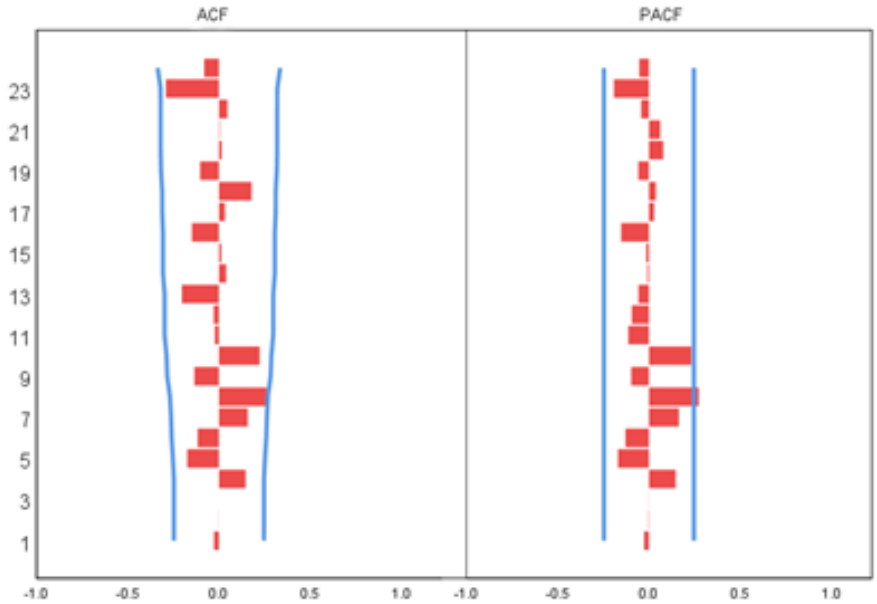


Figure 2. ACF&PACF

Table 2. ARIMA model (0,1,1) test table

Q statistic	Q6(P-value)	0.103(0.748)
	Q12(P-value)	4.234(0.645)
	Q18(P-value)	16.209(0.182)
	Q24(P-value)	25.199(0.120)
	Q30(P-value)	33.923(0.086*)
goodness of fit	R^2	0.856

Note: ***, **, * represent the significance levels of 1%, 5%, and 10% respectively

Through SPSS modeling, we get the ACF and PACF of the residuals, as well as the P value of the stationary R-square and Q-test on the residuals, as shown in Figure 2. From the ACF and PACF graphs of the residuals, it can be seen that the autocorrelation coefficients and partial autocorrelation coefficients of all lag orders are not significantly different from 0. The level of Q6 does not show significance, and the hypothesis that the residual of the model is a white noise sequence cannot be rejected (Table 2). At the same time, the goodness-of-fit R^2 of the model is 0.856, which is close to 1. The performance is excellent, and the model basically meets the requirements. So the ARIMA (0,1,1) model can identify the temperature data well.

Step3: Predicting

Realize the prediction of the global continental temperature in the next 7 years, as shown in the Table 3.

Table 3. ARIMA model forecast temperature

year	Global mean land temperature (°C)
2022	19.407
2023	19.430
2024	19.452
2025	19.475
2026	19.498
2027	19.520

3.2. Neural network model building

For this question, we can also build a BP neural network and GA-BP neural network to predict the global average temperature for 2020-2100. We collected the global average maximum temperature, global average minimum temperature, global total CO2 emissions per year, and global average ocean and land temperatures from 1959-2021 as the four characteristic factors input to the neural network model. Based on the large time span that needs to be predicted in this problem and the insufficient data obtained, we divided the prediction of temperature in 2050 and 2010 into two steps: first, we obtained the input factors for 2022-2050 using the existing data, and then we simulated the global average temperature for 2051-2100 using the input factors for 1959-2050 as the secondary input factors.

3.2.1 BP neural network model preparation

BP neural network is a multilayer feedforward network trained by error back propagation (referred to as error back propagation), which continuously adjusts the weights and bias functions by back propagation of errors and forward propagation of signals to finally generate a very reliable artificial neural network system that simulates the initial problem. Here we use the classical three-layer BP model.

(1) Network initialization

Assume that the number of nodes in the input layer is n, the number of nodes in the hidden layer is l, the number of nodes in the output layer is m. the learning rate is η,and the excitation function is $g(x) = \frac{1}{1+e^{-x}}$ where the excitation function takes the function Sigmoid. For the parameter settings in the BP neural network model, we focus on the maximum number of iterations, error threshold, learning rate, the maximum number of iterations set in this paper is 4000, the error threshold is 1e-6, learning rate 0.01.

(2) Output of the hidden layer:

$$H_j = g(\sum_{i=1}^{n} (w_{ij}x_i + a_j) \tag{2}$$

(3) Output of the output layer:

$$O_k = \sum_{j=1}^{l} (H_j w_{jk} + b_k) \tag{3}$$

(4) Calculation of errors:

$$E = \frac{1}{2}\sum_{j=1}^{l}(Y_k - O_k)^2 \tag{4}$$

where Y_k is the expected output, we note $Y_k - O_k = e_k$ then E can be expressed as $E = \frac{1}{2}\sum_{k=1}^{m} {e_k}^2$

(5) Update of weights:

$$\begin{cases} w_{ij} = w_{ij} + \eta H_j(1 - H_j)x_i \sum_{k=1}^{m} w_{ij} e_k \\ w_{jk} = w_{jk} + \eta H_j e_k \end{cases} \tag{5}$$

This equation shows the process of error back propagation, and our goal is to obtain the smallest error value.

(6) Update of bias:

$$\begin{cases} a_j = a_j + \eta H_j(1 - H_j) \sum_{k=1}^{m} w_{jk} e_k \\ b_k = b_k + \eta e_k \end{cases} \tag{6}$$

3.2.2 GA-BP neural network model preparation

Based on the BP neural network model, there is no corresponding theoretical guidance for the selection of the number of network layers and neurons, and it is easy to fall into local minima, we additionally adopt the improved GA-BP model which based on genetic algorithm optimization. The biggest improvement of this model compared with the original BP is that the optimal parameter values of BP are obtained by genetic algorithm, and then the parameters of BP are initialized accordingly both to avoid the difficulties of unwarranted allocation of layers and number of neurons of BP network [8-10].

The improved part of this model compared to BP neural network model is the embedded genetic algorithm optimization iterations to find out the best parameter values, this paper only introduces the genetic algorithm part of the implementation steps:

(1) Initialize the BP neural network, determine the network structure and learning rules, and the chromosome length of the genetic algorithm.

(2) Initialize the parameters of the genetic algorithm (including the number of iterations, population size, crossover probability and mutation probability, etc.), and the required fitness function for the population selection genetic algorithm.

(3) The choice of select operation is the adaptation scale method, calculated as

$$f_i = \frac{k}{F_i}, P_i = \frac{f_i}{\sum_{j=1}^{N} f_j} \tag{7}$$

where the individual fitness isF_i, k is a coefficient, and N is the overall number of populations.

(4) Processing male parents to generate new population generation by changing the crossover process and genetic algorithm changes

(5) The error is judged whether the accuracy is reached, and if not, the steps (3)(4) are re-executed so that the chromosomes are continuously updated and thus the individuals are constantly updated; when the target is reached, the optimal chromosome is found and assigned to the initial weights and thresholds of the BP neural network.

The above is the specific process of the genetic algorithm, and result we get next will be used as the input of the initial value of the BP neural network. For the genetic algorithm parameter selection, the genetic algorithm parameter settings in the GA optimized BP neural network model mainly include the number of population size, the evolution function parameter, the crossover function parameter, the variation function parameter and the selection function parameter, in turn set to 50,2, [2 50 3], 0.09.

(6) The figure of the algorithm is shown in Figure 3.

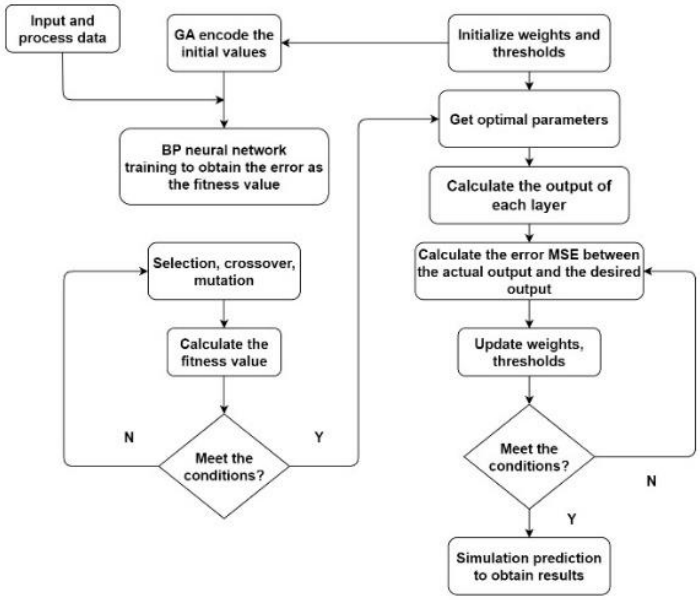


Figure 3. GA-BP neural network algorithm show

3.2.3 BP neural network model and GA-BP network model building

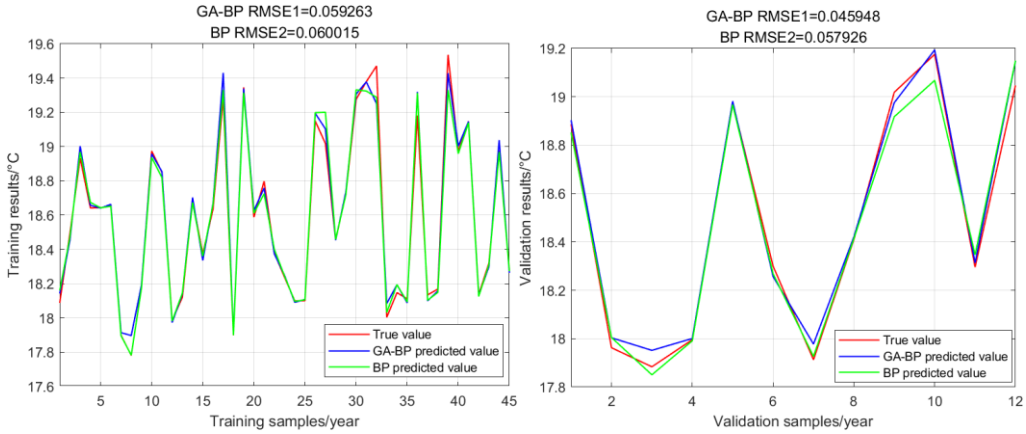


Figure 4. Training set and validation set

To ensure the reliability of the model, we disrupted the input order so that it was not fed into the neural network in increasing year order. The vertical coordinates of our plotted images represent the global average surface temperature and the horizontal coordinates represent the sample counts (Figure 4).

3.2.4 Solution of BP neural network model and GA-BP network model

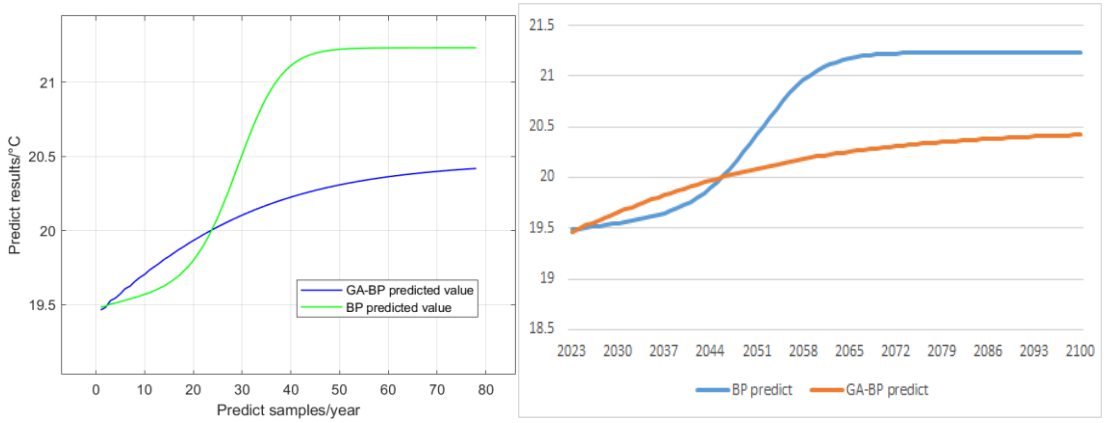


Figure 5. prediction set and Prediction collection format conversion

After the first iteration to obtain the predicted values for 2020-2050, the most recent predicted values are used as new input factors to directly predict the average temperature from 2050-2100. To make the output more intuitive, we take out the data and place it in an Excel sheet to plot it as an image with the correct corresponding year (Figure 5).

3.3. Prediction of temperature based on models

We use the three models of ARMA/BP/GA-BP to predict the temperature respectively, and get the temperature prediction value in the next 80 years. The ARIMA model predicts that the global temperature will reach 20°C in 2050, while the BP model predicts 2046, and the GA-BP model predicts 2045. From the general trend, the temperature has been changing in the past nearly one hundred years, but the overall trend is the upward trend is in line with today's global warming trend, and the three models are consistent on this point.

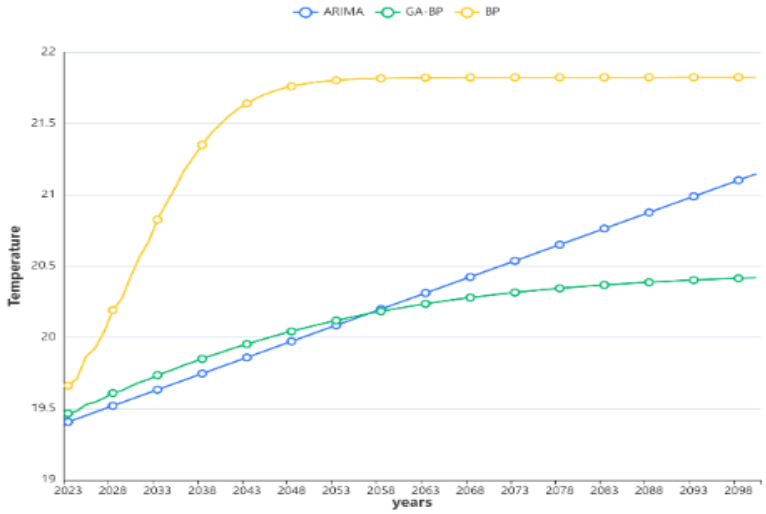


Figure 6. Comparison of three model prediction sets

3.4. Validation of the model

There are many indicators for judging the accuracy of the model. In general scenarios, the three indicators of MAPE, MAE, and MSE can basically meet the analysis requirements, so we use these three indicators measure the accuracy of a predictive model. It can be obtained that the accuracy of GA-BP is higher, so the prediction results of the GA-BP model are more credible (Figures 7 and 8).

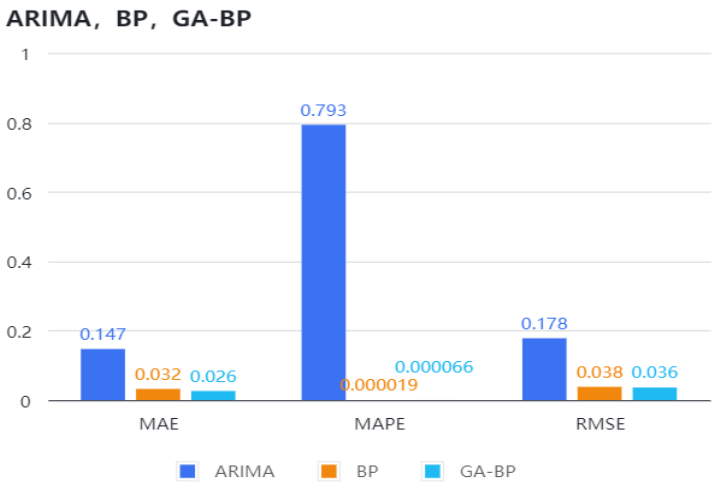


Figure 7. Comparison of the accuracy of the three models

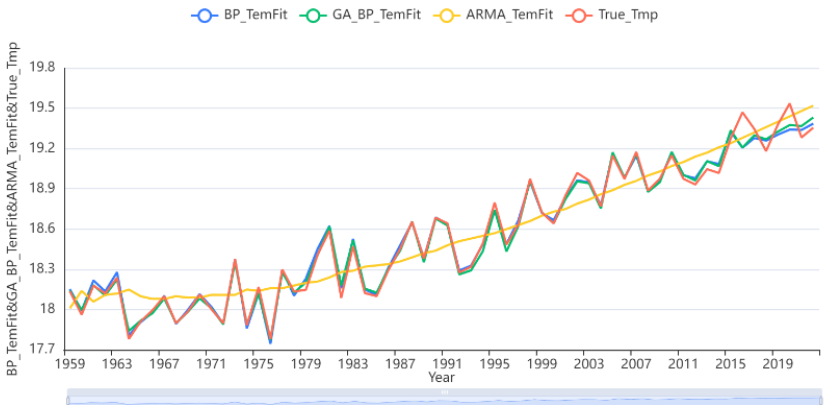


Figure 8. Comparison of the predicted values of the three models

4. Conclusions

We first use the collected 10-year monthly average temperature data and divide them into two groups: 2012.3~2022.2 and 2022.3~2022.10, then conduct the Mann-Whitney U test, and find that the temperature increase in March 2022 did not increase significantly; then we three models are established to predict the future global temperature, the first model is ARIMA, the second model is BP neural network, and the third model is GA-BP neural network. Three models are used to predict the time when the future temperature will reach 20°C, ARIMA predicts the time is 2050, BP is 2046, and GA-BP is 2045. Finally, we use MAE, MAPE, and RMSE to evaluate the accuracy of the model, and find that GA-BP has the highest accuracy, followed by BP, and ARIMA is the worst.

References

[1] Xu Qiang,Dai Linxiu,Shang Ziyin,Zhou Ying,Li Jingyong,Dou Zhi,Yuan Xiaochun,Gao Hui. Application of controlled-release urea to maintain rice yield and mitigate greenhouse gas emissions of rice–crayfish coculture field[J]. Agriculture, Ecosystems and Environment,2023,344.

[2] Takayama Ryota,Nakamura Ryota,Esteban Miguel,Mäll Martin,Ohizumi Kota. Pseudo global warming experiment of flood inundation in the upper White Volta River, Ghana[J]. Journal of Hydrology: Regional Studies,2023,45.

[3] Xiong Jinpeng,Su Ya,He Xueqin,Han Lujia,Huang Guangqun. Effects of functional membrane coverings on carbon and nitrogen evolution during aerobic composting: Insight into the succession of bacterial and fungal communities[J]. Bioresource Technology,2023,369.

[4] Chen Yeyu,Wu Xiaoyun,Lai Jiansheng,Liu Ya,Song Mingjiang,Li Feiyang,Gong Quan. Integrated biochemical, transcriptomic and metabolomic analyses provide insight into heat stress response in Yangtze sturgeon (Acipenser dabryanus)[J]. Ecotoxicology and Environmental Safety,2023,249.

[5] Solís Iris,Álvarez Elena,Barba Emilio. Global warming modifies the seasonal distribution of clutches on a Mediterranean great tit population. [J]. International journal of biometeorology,2022.

[6] Magnoli Susan M,Keller Kane R,Lau Jennifer A. Mutualisms in a warming world: How increased temperatures affect the outcomes of multimutualist interactions.[J]. Ecology,2022.

[7] Decker Mary Beth,Brodeur Richard D.,Ciannelli Lorenzo,Britt Lyle L.,Bond Nicholas A.,DiFiore Bartholomew P.,Hunt George L.. Cyclic variability of eastern Bering Sea jellyfish relates to regional physical conditions[J]. Progress in Oceanography,2023,210.

[8] Qu Huifang,Zhang Weidong,Zhang Guoqing. A novel fault diagnosis scheme for a ballast system[J]. Automatica,2023,148.

[9] Guo Haixia,Cai Rongshuo,Tan Hongjian. Projected harmful algal bloom frequency in the Yangtze River Estuary and adjacent waters[J]. Marine Environmental Research,2023,183.

[10] Sheng Qi,Yi Lingxin,Zhong Bin,Wu Xiaoyu,Liu Liming,Zhang Bin. Shikimic acid biosynthesis in microorganisms: Current status and future direction[J]. Biotechnology Advances,2023,62.