Prediction of La Niña Events Based on BP Neural Network Model

Jie Deng [1, #], Xinfeng Wang [2, *, #], Haodong Liang [3], Siyuan Cai [1], Hao Shi [1], Ruoqing Sun [1]

[1] School of Economics and Management, Tianjin University of Science and Technology, Tianjin, China, 300457

[2] School of Mechanical and Electrical Engineering, Henan University of Science and Technology, Luoyang, China, 471000

[3] School of Computer Information Engineering, Shanxi Technology and Business College, Taiyuan, China, 030000

* Corresponding Author Email: 1822527080@qq.com

[#]These authors contributed equally

Abstract. As an extreme weather phenomenon, La Niña event will cause low water temperature in the equatorial eastern Pacific Ocean and abnormal longitudinal circulation in East Asia, resulting in widespread and continuous abnormal cooling of sea surface temperature in the eastern and central equatorial Pacific Ocean, which will cause frequent precipitation and flooding in the north and high temperature and low rainfall and frequent droughts in the south[1]. Therefore, it is particularly important to assess the damage and propose response strategies[2]. For the main countries involved in the global triple La Niña event, this paper selects four representative countries located in different regions to analyze, through descriptive analysis and graphical visualization, and selects the relevant index data of the last three years: monthly average maximum and minimum temperatures, monthly average precipitation rate, monthly average precipitation days, monthly average wind speed and monthly average water temperature, and finds that New York City and Brazil were affected by the triple La Niña event at the same time and were affected to a higher degree, and the temperature changes in Malasila and Jakara were stable, and the temperature effects brought by the triple La Niña event were smaller[3]. Then, a BP neural network model was established to predict the probability of future La Niña events in each region by modeling pressure, temperature, and salinity data, and it was found that a total of 110 regions would experience La Niña events in the future, with a general probability of 55%-60%..

Keywords: Triple La Niña event, BP Neural Network, Temperature change.

1. Introduction

This summer, climate anomalies occurred in many regions of the world, for example: in China, many regions in the south experienced persistent and rare high temperature weather, and some regions in the north experienced large-scale heavy precipitation weather [4]; in Europe, many cities experienced extremely rare droughts in history [5]. According to the relevant data, these meteorological disasters are not only extremely rare and abnormal weather in decades, but also the most serious meteorological disasters of high temperature, heavy precipitation and drought since human meteorological record data are available [6]. In many cities in southern China and some countries in Europe, the prolonged abnormal high temperatures have led to huge economic losses and a certain degree of casualties; in many cities in northern China, due to the heavy precipitation weather, crop yields have been reduced in large quantities, and in some areas, there are even no harvests [7]. The abnormal high temperature phenomenon and the abnormal heavy precipitation phenomenon were attributed to the triple La Niña event by the meteorological department.

Yet the La Niña event, which has been going on for a long time, is likely to continue until the end of the year or even longer, according to the latest data study by the World Meteorological Organization. This is the first time since the 21st century that a triple La Niña event has occurred,

meaning three consecutive La Niña winters in the Northern Hemisphere. La Niña events are near the equator, where the warm equatorial currents flow from east to west due to the easterly trade winds, causing warmer waters to collect in the western Pacific Ocean, which becomes the warmest water in the world [8]. Conversely, warm seawater from the surface layer of the eastern Pacific Ocean is transported to the western Pacific Ocean and replenished by deeper colder seawater upward, resulting in the phenomenon of cooler seawater in the surface layer of the eastern Pacific Ocean[9]. Research suggests that more La Niña events could have multiple complex effects, potentially causing flooding in Southeast Asia and droughts and wildfires in the southwestern United States, creating multiple hurricane, cyclone and monsoon patterns in the Pacific and Atlantic seas, and triggering weather changes in other regions.

2. Establishment and solution of BP neural network model

2.1. Select Data Metrics

Many regions around the world are affected by triple La Niña events. In order to make the study representative, for the main countries involved in the global triple La Niña events, four countries located in different regions are selected for analysis in this paper, namely, the United States in central North America, Brazil in eastern South America, Malaysia in Southeast Asia, and Indonesia in southeastern Asia.

(1) Maximum and minimum temperatures

The maximum and minimum temperature values are a visual reflection of the climate, too high and too low temperatures are extreme weather effects, so this paper selects the average value of the last three years for each month to represent the maximum and minimum temperature values for that month, and analyzes this data to compare and analyze the impact of different regions of the country.

(2) Average monthly precipitation rate

In this paper, the precipitation amounts of the last three years were selected, and the average values of the precipitation rates of each month in the last three years were observed. Through the analysis, it can be concluded that different regions are affected by the three La Niña events and the degree of impact varies with the time of impact.

(3) Average number of precipitation days per month

In general, the number of precipitation days per month is about 5 days, but it may vary from region to region and from month to month. Paper analyze and observe the impact of triple La Niña events on the precipitation in four countries in these regions to facilitate the subsequent formulation of targeted coping strategies for different countries in these regions.

(4) Monthly average wind speed

The impact of triple La Niña events on different regions will be different, resulting in high temperatures and low wind speeds in some areas and high wind speeds in others, which may also affect the local economic situation.

(5) Monthly average water temperature

The monthly average of water temperature for the last three years is a better visual representation of the monthly temperature of seawater. The occurrence of La Niña is related to the strengthening of the easterly trade winds in the equatorial region, and triple La Niña events can lead to low average water temperatures in some areas, while correspondingly high water temperatures in some areas, making them the warmest in the world.

2.2. Specific analysis of the data

Then we selected the relevant index data for the last three years for these four countries: average monthly maximum and minimum temperature, average monthly precipitation rate, average monthly precipitation days, average monthly wind speed and average monthly water temperature. The data obtained were visualized and analyzed in a line graph to visualize the changes of the relevant indicators for each month in different countries.

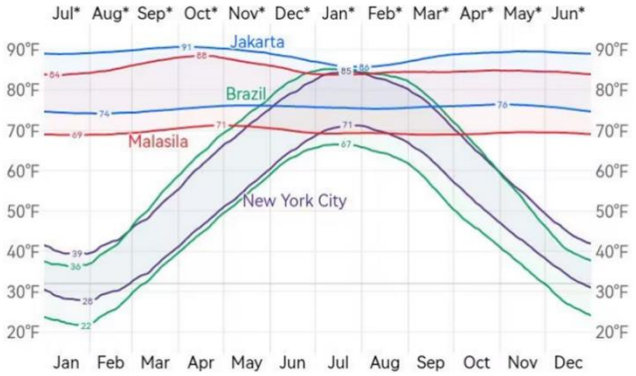


Figure 1. Average monthly maximum and minimum temperatures

As shown in Figure 1., the highest temperature in New York City is 84°F in July and the lowest temperature is 29°F in January, with a maximum temperature difference of 15°F in April and May; the highest temperature in Brazil is 85°F in July and the lowest temperature is 22°F in January, with a maximum temperature difference of 20°F in April and May. The maximum temperature in Malasila is 88 °F in April and the minimum temperature is stable at around 69 °F every month, with a maximum temperature difference of 17 °F in April and May; the maximum temperature in Jakarta is 90 °F in September, October and November and the minimum temperature is 74 °F in July and August, with a maximum temperature difference of 15 °F in July, August and September. The maximum temperature difference is 15 °F in July, August and September.

The maximum and minimum temperatures as well as the maximum temperature difference occurred at basically the same time in New York City and Brazil, indicating that these two regions were affected by the triple La Niña event at the same time, while the minimum temperatures were also lower and the temperature fluctuations were larger, indicating that zhe3 two regions were affected by the triple La Niña to a higher degree. In contrast, the minimum temperature in both Malasila and Jakara is around 70°F, which is not much different from the maximum temperature, indicating that the temperature variation is stable and the temperature impact by the triple La Niña event is less.

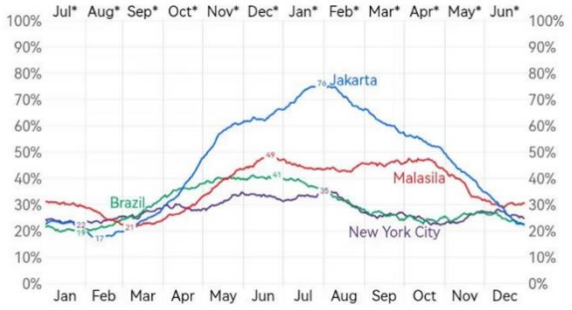


Figure 2. Average monthly precipitation rate

As shown in Figure 2., Jakara has the highest precipitation rate, followed by Malasila, and New York City has the lowest rate.

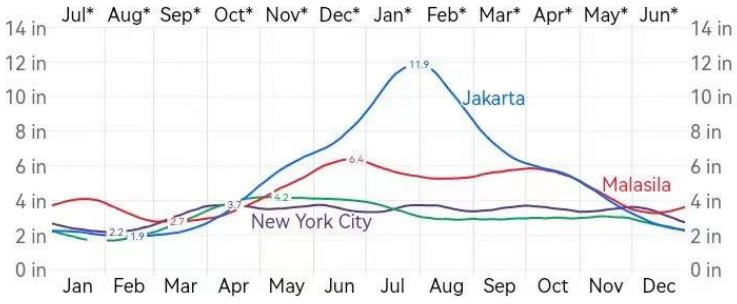


Figure 3. Average number of precipitation days per month

As shown in Figure 3., Jakarta's monthly precipitation starts to drop from January, reaches its lowest in August, and then starts to increase; Malasila's precipitation reaches its highest in June; Brazil and New York City's monthly precipitation is relatively stable.

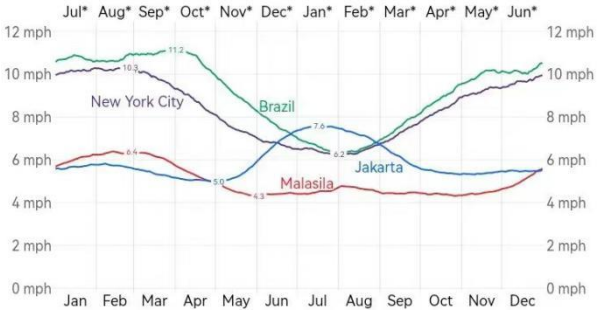


Figure 4. Monthly average wind speed

As shown in Figure 4., under the influence of triple La Niña events, the monthly average wind speed of the four countries in the selected area fluctuates from month to month, with Brazil and New York City fluctuating more, indicating that these two countries are more affected by triple La Niña events.

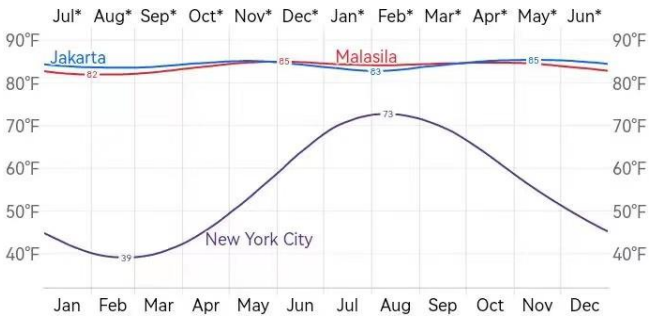


Figure 5. Monthly average water temperature

As shown in Figure 5., New York City's average water temperature varies widely from month to month, reaching a maximum in August and a minimum in February, while the other three countries' monthly average water temperatures do not vary significantly.

2.3. Predicting the likelihood of La Niña events

2.3.1 Problem Analysis

Since there are more missing values in the data such as international temperature in 2022, which is difficult to analyze, the La Niña event occurring in 2017-2018 is selected in this paper to study the modeling. Checking the relevant information, it is known that the specific time of the occurrence of La Niña events in 2017-2018 is starting from October 2017 to April 2018, and the occurrence range is defined as 60 degrees north to 50 degrees south and 80 degrees east to 60 degrees west, so when conducting data processing, a 0-1 variable is established to indicate the occurrence of La Niña events or not, and the data is defined within the above range that is The occurrence of a La Niña event is indicated by "1", otherwise it is not occurring and is indicated by "0".

In this paper, pressure, temperature, and salinity data from November 2016 to March 2019 are selected for analysis, standardized, and then a BP neural network model is established to predict the probability of a possible future La Niña event in each region.

2.3.2 Principle of BP Neural Network

BP neural network is a multi-layer feed-forward neural network with error correction by error back propagation algorithm, and its core feature is that the signal is forward propagation while the error is back propagation [10]. In the forward propagation process, the input signal is processed layer by layer through the input layer and hidden layer, and when it reaches the output layer, if the result does not

reach the desired requirement, it enters the backward propagation process and returns the error signal in the original way to modify the weights of each layer.

The BP neural network contains an input layer, a hidden layer and an output layer [11], in which there can be more than one hidden layer, where the number of nodes in the input and output layers is fixed (the number of variables in the input samples and the number of output labels, respectively), but the number of nodes in the hidden layer is not fixed [12]. Take a BP neural network with a single hidden layer as an example, its network structure is shown in Figure 6.

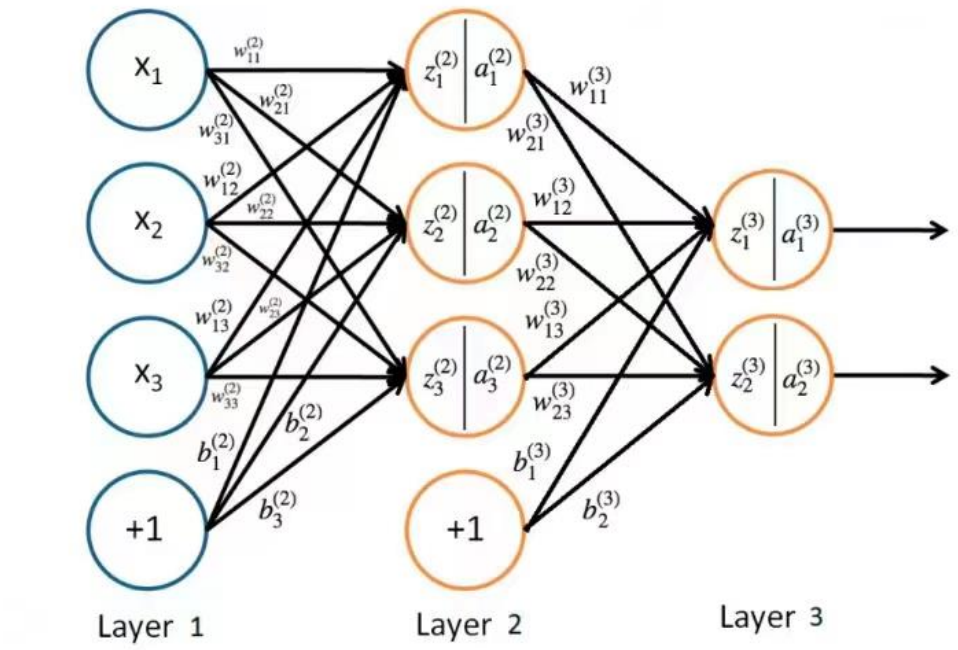


Figure 6. BP neural network structure

In this paper, a BP neural network with a single hidden layer is used, and the principle formulas for each stage are as follows:

1. Forward propagation

Let the input layer have n nodes, the hidden layer have q nodes, the output layer have m nodes, the weight between the input layer and the hidden layer is V, the weight between the hidden layer and the output layer is W, and the input variable is X. Then the outputs of the hidden layer and the output layer are:

$$Z_k = f_1(\sum_{i=1}^{n} V_{ki} X_i), k = 1, 2, \cdots\cdots q \tag{1}$$

$$P_r = f_2(\sum_{i=1}^{q} W_{jk} Z_k), r = 1, 2, \cdots\cdots m \tag{2}$$

2. Error function:

The error function is a function on W, b with different errors using different weights and thresholds.

With p input samples and k outputs, the mean square error function for each input sample is:

$$E(W, b) = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{k}\sum_{j=1}^{k}(\hat{y}_{ij} - y_{ij})^2 \tag{3}$$

where $\hat{y}_{ij}$ is the predicted value of the jth output of the i-th sample and y_{ij} is the corresponding true value.

3. Adjusting weights for errors:

The function is the fastest changing along the direction of the gradient, and the weights are also updated by the gradient descent method in the BP neural network. According to the chain rule and the above relationship, we can get the formulae for the change of weights W and V:

$$\Delta W_{jk} = -\eta \frac{\partial E_p}{\partial W_{jk}} = -\eta(-\sum_{j=1}^{m}(y_{jp} - y'_{jp}))\bullet f'_2(\sum_{k=1}^{q} W_{jk} Z_k)\bullet Z_k \tag{4}$$

$$\Delta V_{ki} = -\eta \frac{\partial E_p}{\partial V_{ki}} = -\eta(-\sum_{j=1}^{m}(y_{jp} - y'_{jp}))\bullet f'_2(\sum_{k=1}^{q} W_{jk} Z_k)\bullet W_{jk} \bullet f'_1(\sum_{i=1}^{n} V_{ki} X_i)\bullet X_i \tag{5}$$

4. BP neural network structure and expressions:

For this problem, we use a three-layer structure, where the implicit layer contains 10 hidden neurons, the implicit layer activation function is tansig, and the output layer activation function is purelin's structure. The structure is shown in Figure 7.

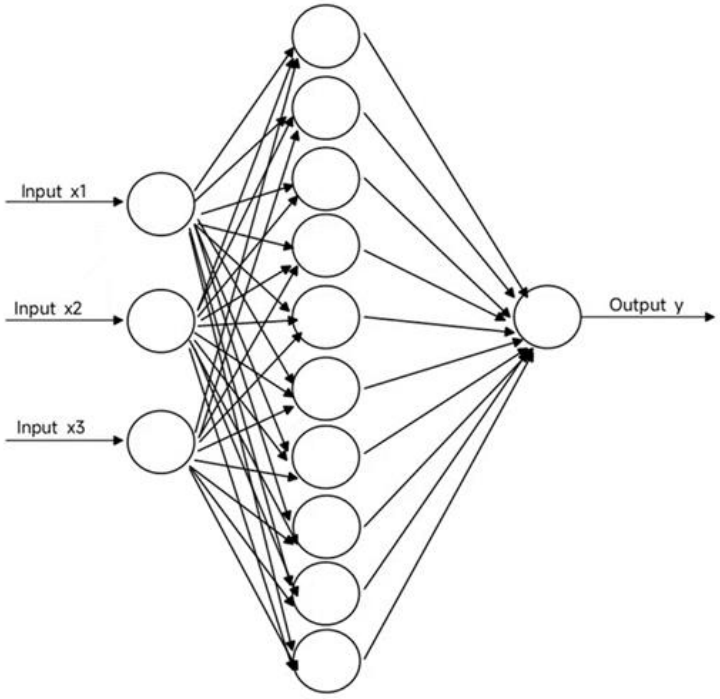


Figure 7. Corresponding network topology

The corresponding mathematical model expression at this point is:

$$\begin{aligned} y = {} & w_{11}^{(2,3)} * \tan sig(w_{11}^{(1,2)} * x_1 + w_{21}^{(1,2)} * x_2 + w_{31}^{(1,2)} * x_3 + b_1^{(2)}) + \cdots + w_{10,1}^{(2,3)} \\ & * \tan sig(w_{1,10}^{(1,2)} * x_1 + w_{2,10}^{(1,2)} * x_2 + w_{3,10}^{(1,2)} * x_3 + b_{10}^{(2)} + w_{1,2}^{(2,3)} \\ & * \tan sig(w_{11}^{(1,2)} * x_1 + w_{21}^{(1,2)} * x_2 + w_{31}^{(1,2)} * x_3 + b_1^{(2)}) + \cdots + w_{10,2}^{(2,3)} \\ & * \tan sig(w_{1,10}^{(1,2)} * x_1 + w_{2,10}^{(1,2)} * x_2 + w_{3,10}^{(1,2)} * x_3 + b_{10}^{(2)} \end{aligned} \tag{6}$$

Where: $w_{11}^{(2,3)}$ represents this weight is the weight between the first node in the second layer and the first node in the third layer

$b_1^{(2)}$represents that this threshold is the threshold of the first node in the second layer.

2.3.3 Model establishment and Analysis

(1) Processing summary

Table.1. Case Summary

		N	percentage
sample	train	20112	69.90%
	test	8664	30.10%
Valid		28776	100.00%
exclude		1440	
total		30216	

As shown in Table.1., the proportion of the training set is about 70% and the proportion of the test set (inspection set) is about 30%.

(2) Network Information

Table.2. Network Information

Input layer	Covariant	1	pressure
		2	temp
		3	salinity
	Number of units		3
	Rescaling method of covariates		standardlized
Hide Layers	Number of units		10
	Activation function		Softmax
output layer	dependent variable	1	Whether the La Nina incident occurred
	Number of units		2
	Activation function		Identity
	error function		Sum of squares

As shown in Table.2., we can see that in this paper, three variables of pressure, temperature and salinity are selected as input layers, an implicit layer is selected which contains 10 nodes, and the final output layer is a 0-1 variable representing whether a La Niña event occurs or not, and the sum of squares method is selected to calculate the error function.

(3) Model Summary and Classification Table

Table.3. Model Summary

train	Sum of squares error	3963.698
	Incorrect forecast percentage	28.20%
	training time	00:00.7
test	Sum of squares error	1743.232a
	Incorrect forecast percentage	29.30%
Dependent variable: whether La Nina event occurs		

Table.4. Model classification table

sample	Measured	predicted		
train		0	1	Correct percentage
	0	14410	42	99.70%
	1	5626	34	0.60%
	Overall percentage	99.60%	0.40%	71.80%
test	0	6117	23	99.60%
	1	2513	11	0.40%
	Overall percentage	99.60%	0.40%	70.70%

As shown in Table.3., we can see that the error rate of the training set is 28.2% and the error rate of the test set is 29.3%; As shown in Table.4., we can see that the correct rate of the training set is 71.8% and the correct rate of the test set is 70.7%. The correct rate is high, therefore, the neural network model has a strong prediction ability for the samples.

(4) ROC curve

The ROC curve is shown in Figure 8., we focus on the area under the curve (also called the area under the curve) with a total area of 1. As shown in Table.5., the areas of 0 and 1 in the model are both greater than 0.6, indicating that the model has good predictive power.

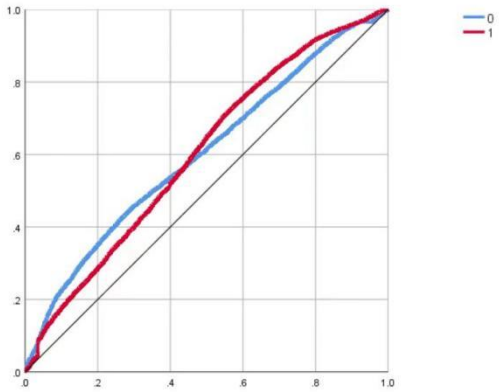


Figure 8. ROC curve

Table.5. Area under the curve

		area
Whether the La Nina incident occurred	0	0.602
	1	0.602

(5) Final model equation

Based on the parameter estimates given in the software, we can obtain the final results of the established BP neural network model as:

$$\begin{aligned} y = {} & 0.476 * \tan sig(-0.244 * x_1 - 0.340 * x_2 + 0.089 * x_3 + 0.328) + \cdots + 0.908 \\ & * \tan sig(-0.220 * x_1 - 0.663 * x_2 + 0.073 * x_3 + 0.257 + 0.524 \\ & * \tan sig(-0.244 * x_1 - 0.340 * x_2 + 0.089 * x_3 + 0.328) + \cdots + 0.092 \\ & * \tan sig(-0.220 * x_1 - 0.663 * x_2 + 0.073 * x_3 + 0.257 \end{aligned} \tag{7}$$

3. Predicted results

Mathematical statistical analysis software was used to obtain partial prediction results, as shown in Table.6. Pressure, temperature and salinity data were used to model the probability of predicting possible future La Niña events in each region.

As shown in Table.6., the last three columns show the predicted probabilities of whether La Niña events will occur in some regions in the future, with the first group being "0" for no occurrence and the second group being "1" for occurrence. The results show that a total of 110 regions will experience a La Niña event in the future, with a general probability of 55%-60%.

Table.6. Partial forecast results

north	east	pressure	temp	psal	present
60.4165	4.3142	575.00	6.08	35.032	0
60.4165	4.3142	600.00	5.15	35.009	0
60.4165	4.3142	625.00	4.74	35.000	0
60.4165	4.3142	643.00	4.52	35.003	0
61.1550	2.2847	550.00	5.08	35.020	0

RBF_PredictedValue	RBF_PseudoProbability_1	RBF_PseudoProbability_2
1	0.427	0.573
1	0.429	0.571
1	0.443	0.557
1	0.451	0.549
1	0.453	0.547

4. Conclusions

This paper establish a BP neural network prediction model based on La Niña event to predict the probability of occurrence in different regions, so as to take preventive measures in advance to reduce the huge losses caused by La Niña event. The model can also be used to predict the occurrence probability and damage assessment of various adverse events caused by other natural disasters and man-made causes, and the modeling process is not complicated and has strong practical and realistic significance.

References

[1] Gao Chuan, Chen Maonan, Zhou Lu, Feng Licheng, Zhang Ronghua. Evolution of persistent double La Niña events in the tropical Pacific Ocean from 2020 to 2021 [J]. China Science: Earth Science,2022,52(12):2353-2372.

[2] Qiu Hui. The "triple" La Niña event is coming, will this winter be colder? [J]. China Report, 2022(12):86-88.

[3] Zhou Mo, Fang Xingnan, Yu Wei, Liu Lianwei. Spatial and temporal differences in the habitat distribution of softfish in the northwest Pacific Ocean under El Niño and La Niña events [J]. Journal of Shanghai Ocean University,2022,31(04):984-993.

[4] Wen Jian, Zhang Yanjing, Li Yikai, Li Yisong, Yu Wei. Spatial and temporal variability of mackerel habitat in the East China Sea under different intensity La Niña events [J]. China Fisheries Science,2022,29(06):903-913.

[5] Zhou Mo, Fang Xingnan, Yu Wei, Liu Lianwei. Differences in spatial and temporal distribution of softfish habitat in the northwest Pacific Ocean under El Niño and La Niña events [J]. Journal of Shanghai Ocean University,2022,31(04):984-993.

[6] Chen Mingcheng, Huang Yvrong. Analysis of the causes of the difference between the life history evolution of the typical La Niña event and the 1988-1989 La Niña event [J]. Journal of Atmospheric Sciences,2021,44(03):428-440.

[7] Guan Wenjiang, Chen Xinjun, Li Gang. Effects of sea surface water temperature and La Niña events on the spatial and temporal variability of mackerel resources in the East China Sea [J]. Journal of Shanghai Ocean University,2011,20(01):102-107.

[8] Hu Xinyue, Hu Wenyun. Impact of El Niño/La Niña events on the climate of Xuancheng [J]. Anhui Agronomy Bulletin,2018,24(22):139-141+164.

[9] Qi Yanwen, Zhao Meilan, Ma Qiang. Analysis of the causes of climate forecast errors in autumn and winter of Tongliao City in 2020 under the background of La Niña [J]. Inner Mongolia Meteorology,2021(05):23-26.

[10] Wang Shunan, Gu Fengfeng, Li Junhua, Zhao Dezhao. BP neural network-based prediction of deep-water channel silt return in the Yangtze estuary[J]. People's Yangtze River,2023,54(02):206-213.

[11] Wang Deming, Wang Li. Short-term wind speed prediction model based on genetic BP neural network[J]. Journal of Zhejiang University (Engineering Edition),2012,46(05):837-841+904.

[12] Li Zhen, Wei Yong. Research on infrared temperature measurement compensation algorithm based on BP neural network[J]. Machine Manufacturing and Automation,2023,52(01):170-172+176.