Problem Chosen	2022 ShuWei Cup Summary Sheet	Team Control Number
D		2022111414030

Summary

At present, people are constantly controversial about the Triple La Niña event. On the one hand, due to the continuous accumulation of the greenhouse effect, the energy absorbed and emitted by the earth-atmosphere system is not balanced, resulting in the continuous accumulation of energy in the earth-atmosphere system, resulting in global warming ; on the other hand, in recent years, extremely cold weather has occurred frequently. The La Niña event caused by the continuous abnormal cold sea surface temperature in the eastern and central equatorial Pacific has brought many impacts on most regions of the world. According to the latest data of the World Meteorological Organization, the La Niña event that has lasted for a long time is likely to continue until the end of this year or longer[1]. This will be the first ' triple ' La Niña event in the 21 st century, which will have a great impact on human life. Therefore, it is necessary to analyze the global climate data over the years by means of computer to find the law of La Niña phenomenon, so as to deepen people 's understanding of the Triple La Niña event.

Aiming at the first question: We counted the major countries and regions involved in the global Triple La Niña event and analyzed the impact on these regions. Then we use Surfer software and Kriging interpolation method to model and interpolate the temperature field[2]. It is found that the global temperature is increasing with time in the past 100 years. The wavelet analysis method is used to analyze the variation of ocean surface temperature. We predict that the Triple La Niña event is very likely to occur in the future and will continue until December 2022 to February 2023.

Aiming at the second question: Taking China as an example, we selected the relevant data of forest land area in southern China affected by heat and drought disasters through comparative analysis methods, discarded some redundant variables through previous data preprocessing, and reduced the number of variables according to the obtained Pearson correlation coefficient[3-5]. The importance of the processed data was ranked by grey correlation degree and random forest model, and the first 9 variables that had significant influence on the area of drought forest land were selected. The effect of triple La Nina phenomenon on the loss of heat and drought disaster was quantitatively evaluated, and the effective coping strategies were put forward.

Aiming at the third question: Taking Australia as an example, the comparative method is used to select the data of industrial economic growth rate as the appropriate index, and the loss caused by flood is quantitatively evaluated by establishing the influence model of triple La Niña phenomenon on the index. We statistically analyze the growth rate of agricultural income, GDP and trade data of Australia from 2010 to 2021, use the regression prediction model to predict the corresponding indicators for 2022, and compare the predicted results with the actual statistical results in 2022, so as to quantitatively evaluate and analyze the various disaster losses caused by the floods under the Triple La Niña event, and propose effective response strategies[6].

Aiming at the fourth question: Based on the contents of the first three questions, we wrote a report to the National Meteorological Administration, elaborated our understanding of the Triple La Niña event and gave some relevant countermeasures.

Keyword: Triple La Niña; Wavelet analysis; Grey correlation; Random forests; Regression prediction

Content

1. Introduction

1.1 Background

Floods and heatwaves are currently causing climate disruptions across the globe. As the global climate warms, the frequency of extreme weather events is changing, showing an increasing trend. In fact, humans are adapting to extreme weather events, with hot summer heat, violent hurricanes, biting cold and torrential floods becoming "regular" events, and good weather is now considered a luxury. According to the Hong Kong Observatory, the frequency of extreme weather events caused by global warming is expected to increase worldwide in 2013.

In recent years, disaster losses due to extreme weather have been increasing, how to make accurate climate predictions of extreme events is still an urgent issue and models for prediction and prevention need to be improved. An article in the British journal 《Nature》 in June warned that more La Niña events would have multiple impacts, such as increasing the probability of flooding in Southeast Asia[7], increasing the risk of drought and mountain fires in the southwestern United States, and creating multiple hurricane, cyclone and monsoon patterns in the Pacific and Atlantic Oceans, as well as triggering weather changes in other regions.

Our task is therefore to develop a simple, intuitive and effective model of the triple La Niña environment, using available statistical data, to understand the factors affecting La Niña and the impact of disasters brought about by La Niña, and to carry out a quantitative analysis, which can evaluate the many factors affecting the Triple La Niña imaginary, such as temperature. Based on this, the model will be replicated and analyzed worldwide and applied to areas of the world where forecasting is needed to raise awareness of disaster preparedness and to monitor policy makers' policies on disaster preparedness projects.

1.2 Problem analysis

Question one: This problem first determines the main countries and regions affected by the Triple La Niña, analyzes the impact of the local Triple La Niña event, and obtains the corresponding climate change characteristics. Then we need to collect the annual change data of ocean surface temperature for analysis, and use wavelet analysis method to predict the possibility of future triple La Nina events.

Question two: This problem needs to first find the relevant data of the impact of heat and drought disasters in a country through comparative analysis methods, select appropriate indicators to establish an evaluation model of the impact of the Triple La Niña phenomenon on these indicators, and then use the gray correlation degree and random forest model to rank the importance of these indicators. The first nine variables that have a significant impact on the area of dry forest land are selected to quantitatively evaluate the impact of the Triple La Niña phenomenon on the loss of heat and drought disasters. Finally, targeted measures are given for the impact.

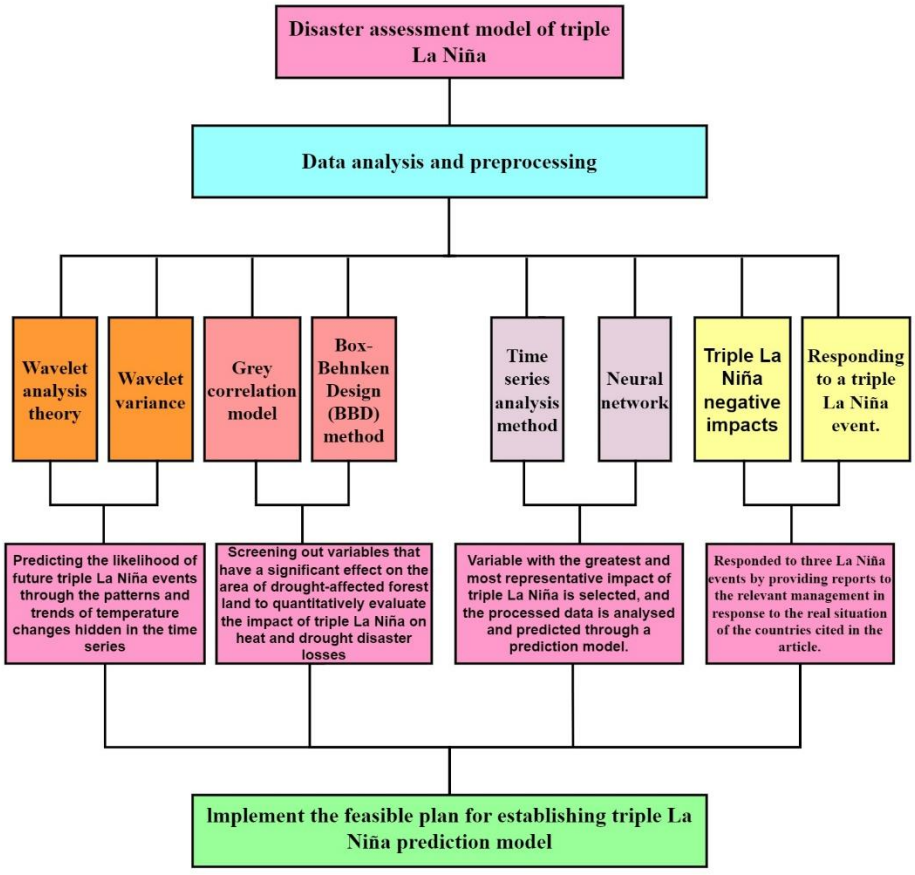
Question three: This problem needs to investigate the various disaster losses caused by a country 's floods under the Triple La Niña event, such as the growth rate of agricultural income, GDP, trade data from 2010 to 2021, etc., and then use the regression prediction model to predict the corresponding indicators of the country in the next few years. The predicted results are compared with the actual statistical results of the year, and the various disaster losses caused by the floods under the Triple La Niña event can be obtained. Finally, targeted measures are given for the impact.

Question four: The solution to this problem is based on the first three issues. By summarizing the main countries and regions affected by the Triple La Niña event, the main impact of the Triple La Niña event is analyzed, and the possibility of the Triple La Niña event in the future is predicted. Based on these problems, write a report to the relevant management to provide reasonable suggestions for dealing with the complex Triple La Niña event.

1.3 Model assumes

1. Assuming that all official data are true and valid;
2. Assuming that all model evaluation and prediction results do not consider the influence of human intervention factors.

1.4 Technical route



2. Problem 1 model establishment and solution

2.1 Analysis of problem 1

Question 1 requires statistical analysis of the major countries and regions involved in the global Triple La Niña event to predict the possibility of future Triple La Niña event. For this problem, we must first determine the main countries and regions affected by the Triple La Niña event, analyze the local impact of the Triple La Niña event, and obtain the corresponding climate change characteristics. For the prediction of the possibility of the future occurrence of the Triple La Niña event, it is necessary to collect the annual change data of the ocean surface temperature for analysis. Because the change law of the ocean surface temperature is affected by many factors, it is a typical non-stationary sequence. It not only has the characteristics of periodicity and trend, but also has the characteristics of mutation, randomness and "multi-time scale" structure, which contains multi-level evolution characteristics for time domain and frequency domain analysis. Therefore, the wavelet analysis method with time-frequency multi-resolution function can be selected to obtain the temperature change law and trend hidden in the time series. Predict the possibility of future Triple La Niña event.

2.2 Establishment of Wavelet Analysis Model

Wavelet analysis theory has been widely used in many nonlinear science fields such as signal analysis, image processing, pattern recognition, seismic exploration and atmospheric science. In the study of time series, this method is mainly used for noise reduction and filtering of time series, obtaining information coefficient and fractal dimension, monitoring break points, identifying change cycle and analyzing multi-time scale.

2.2.1 Establishment of wavelet function

The basic idea of wavelet analysis is to represent or approximate a signal or function by a series of wavelet functions. Therefore, this function is the basis of wavelet analysis. This function refers to a class of functions that are oscillating and can quickly decay to zero, that is, wavelet function : $\psi(t) \in L^2(R)$， And satisfy :

$$\int_{-\infty}^{+\infty} \psi(t)dt = 0 \qquad (2.1)$$

In the formula，$\psi(t)$ is the base wavelet function, which can form a family of functions by scaling and shifting the time axis :

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right) \qquad (2.2)$$

In the formula，$a,b \in R, a \neq 0$, a is the scale factor which reflects the period length of wavelet, b is the shift factor which reflects the shift of reaction time, $\psi_{a,b}(t)$ is a sub-wavelet. In this problem, the basis wavelet function is judged by comparing the

error between the results obtained by different wavelet analysis and processing signals and the theoretical results, and the required basis wavelet function is selected.

2.2.2 Wavelet transform

If $\psi_{a,b}(t)$ is a wavelet given by (2.2), for a specific energy limited signal $f(t) \in L^2(R)$, its Continue Wavelet Transform (CWT) is :

$$W_f(a,b) = |a|^{-1/2} \int_{-\infty}^{+\infty} f(t)\psi\left(\frac{t-b}{a}\right)dt \qquad (2.3)$$

In the formula， $W_f(a,b)$ is the wavelet transform coefficient, $f(t)$ represents a signal or square integrable function, a is the scaling, b is the translation parameter ; $\overline{\psi}\left(\frac{t-b}{a}\right)$ is the complex conjugate function of $\psi\left(\frac{t-b}{a}\right)$. In this problem, since the obtained temperature data points are discrete points, the function $f(h\Delta t),\ (\mathrm{h}=1,2,\ldots,\mathrm{N})$, Δt is the time interval, and the unit is generally days or months.

The discrete wavelet transform of formula (2.3) can be expressed as follows :

$$W_f(a,b) = |a|^{-1/2}\ \Delta t \sum_{k=1}^{N} f(k\Delta t)\overline{\psi}\left(\frac{k\Delta t - b}{a}\right) \qquad (2.4)$$

The basic principle of wavelet analysis can be known from Formula (2.3) or (2.4), that is, the frequency information of the signal is obtained by changing the stretching scale a, and then the general picture or detail of the signal is analyzed, so as to realize the analysis of different time scales and spatial local characteristics of the signal.

2.2.3 Wavelet variance

By integrating the average value of the wavelet coefficients in the b domain, the wavelet variance is obtained, that is :

$$Var(a) = \int_{-\infty}^{+\infty} \left|W_f(a,b)\right|^2 db \qquad (2.5)$$

The variation curve of wavelet variance with scale a is called wavelet variance graph. It can be seen from (2.5) that it can reflect the distribution of signal fluctuation energy with time scale a. Therefore, the wavelet variance diagram can be used to determine the relative intensity of different scale disturbances in the temperature data volume and the main time scale, namely the main period, which is the law contained in the historical data of the ocean surface temperature.

2.3 Problem one Solution

2.3.1 Regional impact

The survey identifies the major countries and regions involved in the global Triple La Niña event, as shown in Fig. 2.1.

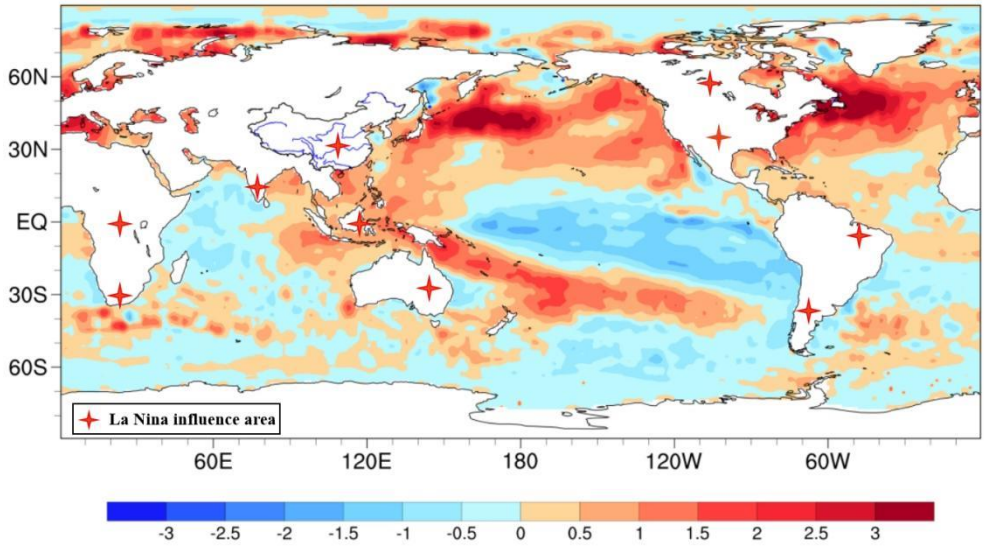


Fig. 2.1 Distribution map of triple La Nina influence area

The main impact areas in the above figure basically cover all climate regions in the world, and the Triple La Niña event has affected the climate in many parts of the world. In China, there are so-called droughts in the south and floods in the north. Indonesia and eastern Australia have increased precipitation. Northeast Brazil, India and southern Africa are prone to floods. The precipitation near the coast of South America has decreased. Droughts often occur in central Africa and southern America. Typhoons are active in the South China Sea and the western Pacific. During the period of La Nina 's influence, the risk indexes of yield reduction of Canadian wheat, American corn and Argentine soybean were high, and the risk of yield reduction of corn, wheat and rice in China was also high. It is worth mentioning that, superimposed on the background of global warming, La Niña climate impact characteristics are also changing.

The variation of ocean surface temperature is analyzed by the wavelet analysis method in the 2.2 summary.

2.3.2 The variation law of ocean surface temperature

This law is solved and analyzed by wavelet analysis model. The specific analysis process is as follows.

(1) Wavelet data expansion

For this time series problem, both ends of the time series may have an impact on the results, that is, the so-called “ boundary effect ”. In order to eliminate this effect, it is necessary to extend the temperature data. After wavelet transform, the extended data is deleted to restore the original number and appearance of the data. In Fig. 2.2, the red box represents the original data, and the brown box represents the extended data. According to the data, the original time series extend 31 and 32 to left and right.

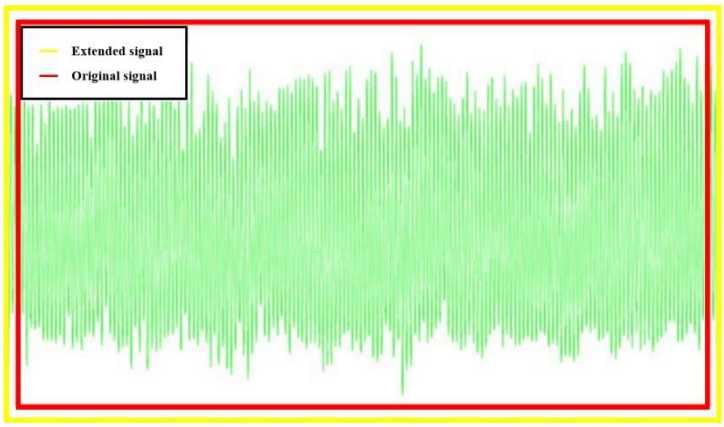


Fig. 2.2 Wavelet data expansion diagram

(2) Drawing wavelet coefficients real part contour map

The contour map of the real part of the temperature wavelet coefficient is shown in Fig. 2.3.

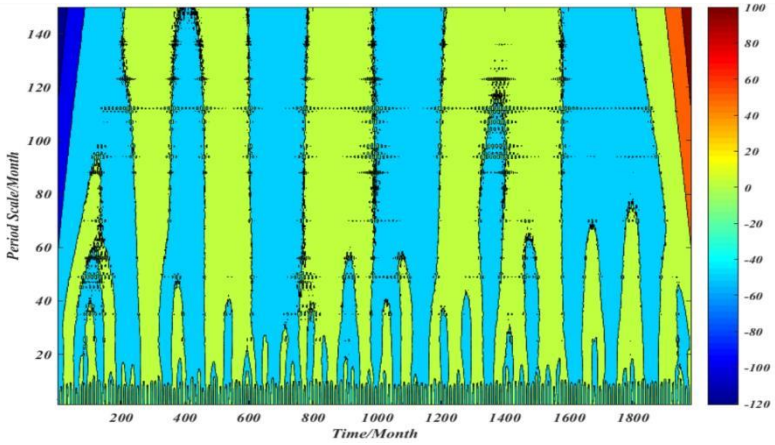


Fig. 2.3 Wavelet coefficient real part contour map

From Fig. 2.3, it can be seen that there are multiple oscillation periods in the evolution of ocean temperature in 165 years. There is a small center and a large center on the whole time scale, which correspond to the beginning and end of the sequence, respectively. It can be explained that the ocean temperature as a whole shows an upward trend, and the average temperature is the largest.

(3) Draw wavelet coefficient modulus and contour map

First calculate the modulus and modulus of wavelet coefficients, and then draw modulus and modulus contour map. As shown in Fig. 2.4.

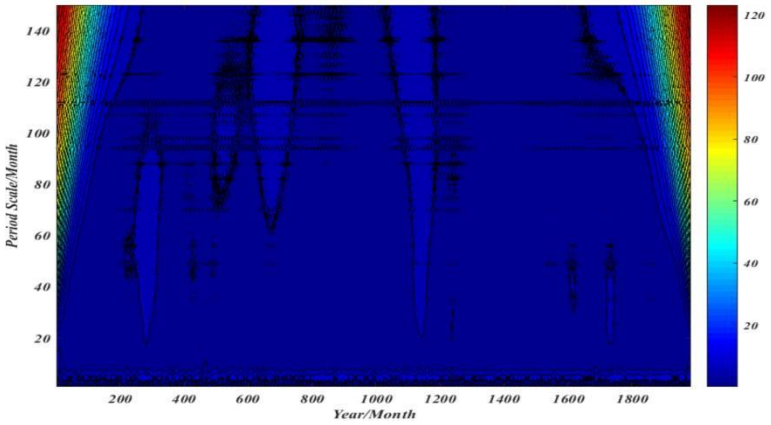


Fig. 2.4 Wavelet coefficient modulus and contour map

(4) Draw wavelet variance diagram

Wavelet variance diagram can reflect the distribution of fluctuation energy of temperature time series with scale, which can be used to determine the main period in the process of temperature change, as shown in Fig. 2.5.

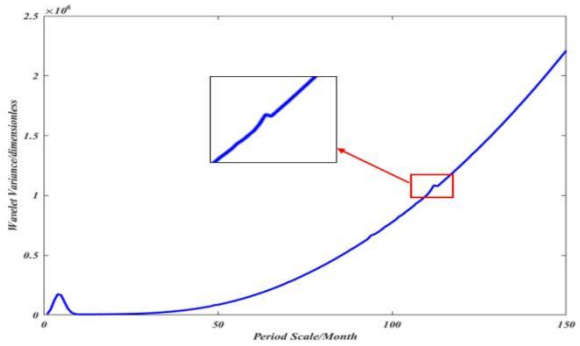


Fig. 2.5 wavelet variance diagram

According to Fig. 2.5, there are two peaks in the wavelet variance diagram of the time series, which correspond to 6-month time scale and 112-month time scale respectively. Among them, the peak value of the 6-month time scale is the largest, indicating that the periodic oscillation at this time scale is the strongest, which is the largest main period of temperature evolution in 165 years. The peak value of 112 months time scale is small, corresponding to the second main period of temperature change. Therefore, the fluctuation of 6-month time scale and 112-month time scale controls the variation characteristics of ocean surface temperature in the whole time domain.

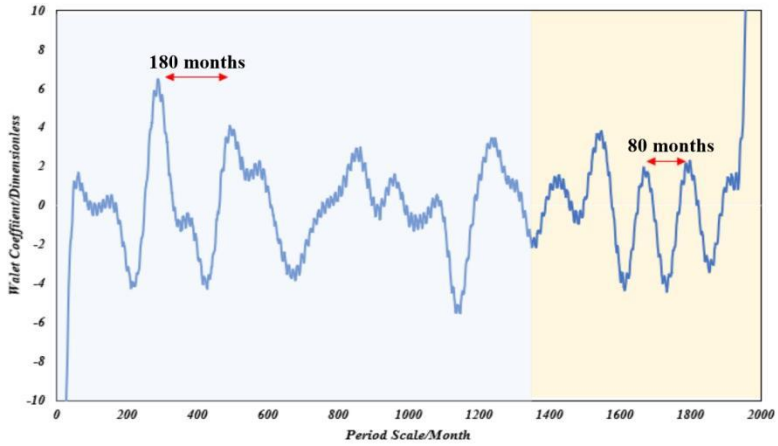


Fig. 2.6 First main cycle trend chart

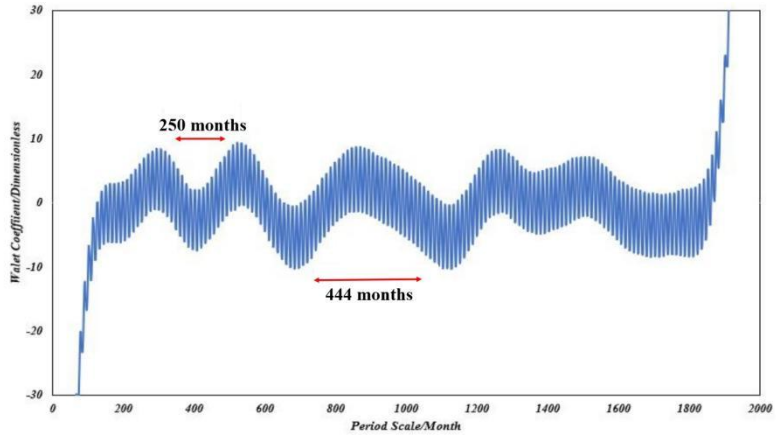


Fig. 2.7 Second main cycle trend chart

In the first main cycle trend diagram (Fig. 2.6), we can see that in the first main cycle of the 6-month time scale, the previous average period is about 180 months, indicating that the ocean indicates that the temperature is in a relatively stable state, mainly affected by solar radiation, atmospheric heat dissipation and other factors, and the average period is shortened to 80 months in the later period, indicating that the ocean shows that the temperature change tends to be unstable in recent years, which may be related to human factors such as greenhouse gas emissions.

On the second main cycle of the 112-month time scale (Fig. 2.7), the average period varies from 250 months to 440 months, and the contribution rate of this part is small. We believe that this part of the uncertain periodic activity may be related to natural factors such as the Pacific Decadal Oscillation and the La Niña event. According to Minobe (1999), the oscillation period of the Pacific Decadal Oscillation is about 10-20 years, even up to 50-70 years, which is consistent with the period range of the second main cycle.

According to the survey data of the National Meteorological Center, the changes of global ocean surface temperature (SST) in 1900,2000,2010,2020 and 2021 are plotted, as shown in Fig. 2.8-Fig. 2.12.

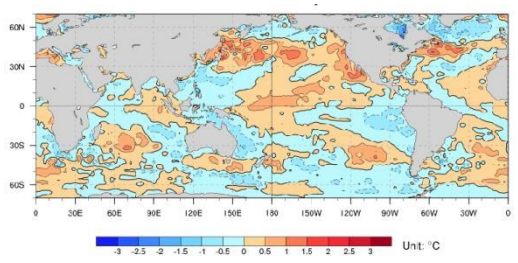


Fig. 2.8 Distribution of Global Sea Surface Temperature Anomalies in 1990

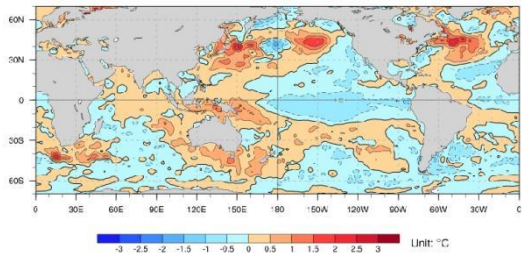


Fig. 2.9 Distribution of Global Sea Surface Temperature Anomalies in 2000

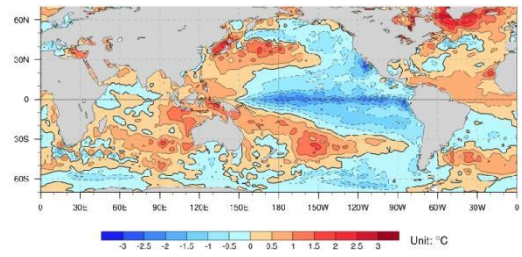


Fig. 2.10 Distribution of Global Sea Surface Temperature Anomalies in 2010

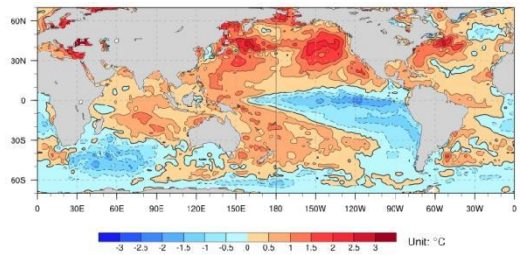


Fig. 2.11 Distribution of Global Sea Surface Temperature Anomalies in 2020

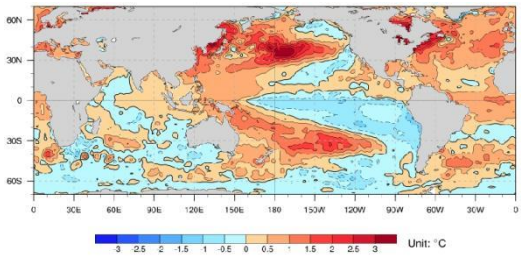


Fig. 2.11 Distribution of Global Sea Surface Temperature Anomalies in 2010

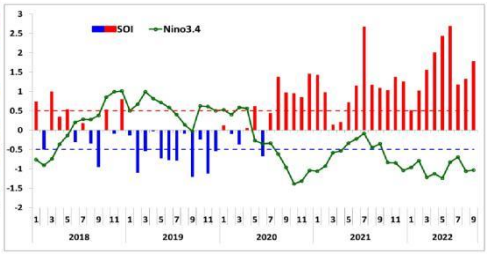


Fig. 2.12 The monthly evolution of sea temperature index (℃) and SOI index in Niño3.4 region

According to the changes of global sea surface temperature anomalies in 1900, 2000, 2010 , 2020 and 2021, the following rules can be summarized :

① The temperature in the Indian Ocean and Southeast Asia generally showed a gradual upward trend, rising from about -0.5 ℃ in 1900 to more than 0.5 ℃ in 2010. As time goes on, the inhomogeneity of temperature distribution gradually diminishes, and by 2020 the sea surface temperature in the region will be essentially the same.

② East Pacific, Atlantic Ocean region sea surface temperature increased, the overall increase is slightly lower than the Indian Ocean region.

③ The temperature of the equatorial Pacific shows a general trend of rising-falling-rising, indicating that the temperature is unstable in the region, which may be related to the La Niña phenomenon in the region and destroy the temperature cycle of the ocean itself.

④ The isothermal line gradually moves closer to the Antarctic and Arctic, indicating that the temperature in the Antarctic and Arctic is also increasing. But the overall trend is slightly lower than the above regions.

⑤ The temperature change in the Northern Hemisphere and the equatorial region is greater than that in the Southern Hemisphere. After more than 30 years of temperature evolution, the overall temperature in the Southern Hemisphere has not changed much.

2.4 Summary of question 1

Through the statistical analysis of the global Triple La Niña event impact area, we have clarified the main impact scope and degree of the Triple La Niña event, and obtained the corresponding climate change characteristics. The main impact areas of the Triple La Niña event are distributed in Australia, the United States, Canada, China, northeastern Brazil, India and southern Africa, and the coast of South America. It will bring natural disasters such as floods, droughts, and precipitation changes to the local area, which will affect the local economic production.

Through the establishment of wavelet analysis model, it is concluded that there are multiple oscillation periods in the ocean temperature time series, and the overall trend is upward. The maximum main period of the time series is 6 months, the second main period is 112 months, and the oscillation is the strongest under the maximum main period. In the first main cycle trend diagram of the six-month time scale, the average cycle is shortened from 180 months in the early stage to 80 months in the later stage, indicating that the ocean temperature changes from a relatively stable state to an unstable state. This phenomenon may be related to human factors such as greenhouse gas emissions. In the second main cycle of the 112-month time scale, the average period ranges from 250 months to 440 months, and the contribution rate is small. This phenomenon may be related to natural factors such as uncertain periodic activity, Pacific interdecadal oscillation, La Niña phenomenon, and the period range of the second main cycle is consistent with the Pacific interdecadal oscillation period obtained from the survey, which further verifies the correctness of the speculation.

Using the ocean temperature data, the spatial distribution map of ocean temperature at different times is drawn, and it is concluded that the temperature in most parts of the world is on the rise. Among them, the temperature rise in the southeast Asia of the Indian Ocean is the largest. The temperature change in the northern hemisphere and the equatorial region is significantly stronger than that in the southern hemisphere. The temperature in the equatorial Pacific region shows an unstable change and there is a low temperature region, which is closely related to the Triple La Niña event. In the future, the Triple La Niña event is very likely to occur and continue until December 2022 to February 2023.

3. Problem 2 model establishment and solution

3.1 Analysis of problem 2

For the second question, the topic requires to take a country as an example to evaluate and analyze the losses of various heat and drought disasters under the Triple

La Niña event, and propose targeted coping strategies, that is, first find the relevant data of the impact of heat and drought disasters in the region through comparative analysis methods, and discard some redundant variables through pre-data preprocessing. According to the obtained Pearson correlation coefficient, the number of variables is reduced, and appropriate indicators are selected to establish an evaluation model for the impact of the Triple La Niña phenomenon on these indicators, so as to quantitatively evaluate and analyze the losses of heat and drought disasters. In order to select the index with great influence of variables on heat and drought, the importance of the processed data is sorted by grey correlation degree and random forest model, and the first 9 variables with significant influence on the area of dry forest land are selected to quantitatively evaluate the influence of Triple La Niña phenomenon on heat and drought disaster loss. And can give targeted measures for the impact.

3.2 Establishment of grey correlation model

The degree of gray correlation can be used to measure the closeness between factors, that is, correlation. The greater the degree of gray correlation, the closer the two are. The measurement value of gray correlation degree can clearly analyze the relationship and closeness between various factors, so it can also be used to determine the optimal factor. The specific steps of the establishment process based on the grey relational analysis method are as follows:

(1) Determine the reference series (evaluation criteria)

Suppose the corresponding year of the evaluation object is m years, so the reference series can be expressed as:

$$x_0 = \{x_0(k) | k = 1, 2...m\} \tag{3.1}$$

There are n factors to be evaluated, and the comparison series can be expressed as:

$$x_i = \{x_i(k) | k = 1, 2...5\}, i = 1, 2, ..., n \tag{3.2}$$

(2) Calculate the degree of gray correlation

$$\xi_i(k) = \frac{\min\limits_s \min\limits_t |x_0(t) - x_s(t)| + \rho \max\limits_s \max\limits_t |x_0(t) - x_s(t)|}{|x_0(k) - x_i(k)| + \rho \max\limits_s \max\limits_t |x_0(t) - x_s(t)|} \tag{3.3}$$

In order to compare the correlation coefficient of the sequence x_i to the reference sequence x_0 on the k index, $\rho \in [0,1]$ is the resolution coefficient. Among them, $\min\limits_s \min\limits_t |x_0(t) - x_s(t)|$ and $\rho \max\limits_s \max\limits_t |x_0(t) - x_s(t)|$ are the two-level minimum difference and the two-level maximum difference, respectively. Generally speaking, the larger the resolution coefficient ρ, the larger the resolution; the smaller the ρ, the smaller the resolution.

(3) Calculate the grey weighted correlation degree

The calculation formula of gray weighted correlation degree is:

$$r_i = \sum_{k-1}^{n} w_i \xi_i(k) \tag{3.4}$$

In the formula, w_i is the weight corresponding to each index. $\omega = [\omega_1, \omega_2, ...\omega_n]$,

where $\omega_k\ (k=1,2,...n)$ is the weight corresponding to the k evaluation index. r_i is the gray weighted correlation degree of the i evaluation object to the ideal object.

(4) Evaluation and analysis

According to the degree of gray weighted relevance, sorting the evaluation objects can establish the relevance order of the evaluation objects. The greater the relevance degree, the better the evaluation effect.

3.3 Grey correlation model solution and analysis

After data preprocessing, the integrity rate of the quantitative variable data analyzed is relatively complete, up to more than 90 %, and the data integrity rate is high, which increases the credibility of the model solution results.

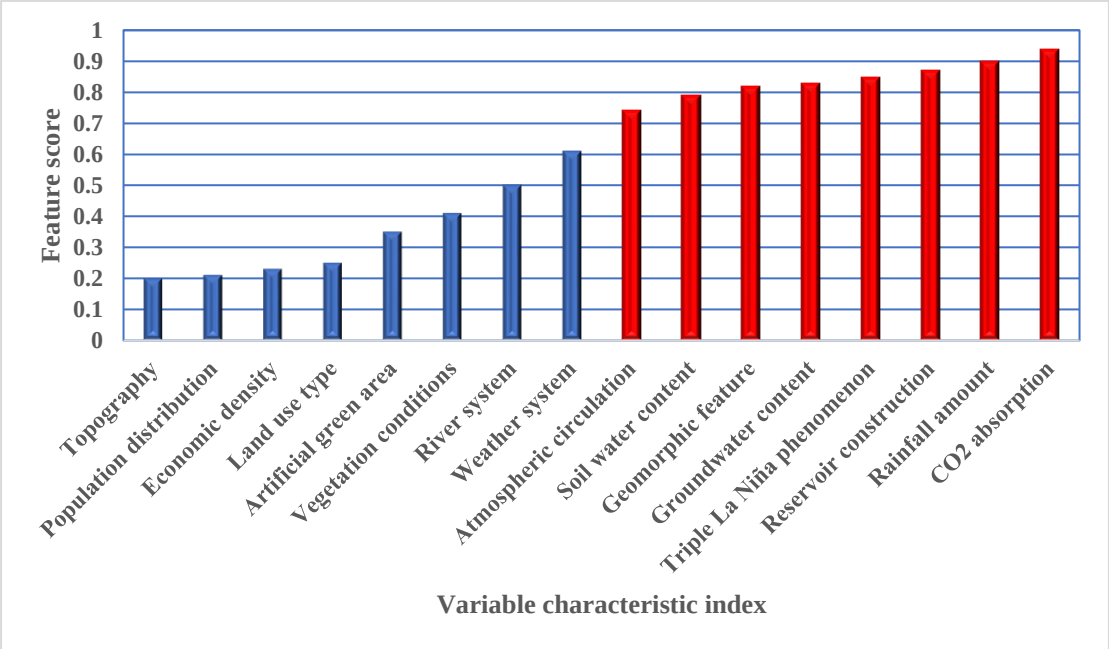


Fig. 3.1 Select the correlation coefficient comparison diagram of variables

It can be seen from Fig. 3.1 that the correlation degree of the 16 selected variables (climate warming, rainfall, Triple La Niña phenomenon, reservoir construction, etc.) is significantly greater than that of the remaining variables (population distribution, economic density, artificial greening area, etc.). It can be concluded that the accuracy of the variables selected in this model is high[8].

3.4 Establishment of random forest model

(1) Given sample set T, containing N samples, denoted as $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$,each sample has M attributes, denoted as $f_1, f_2, \dots, f_M$, respectively, represent the characteristic parameters of the variables to be analyzed that have an impact on biological activity; given an integer m, it is used to indicate tree

generation, and m attributes are selected as candidate attributes; given an integer t, it is used to indicate trees in the forest Number of.

(2) With the replacement method, n samples are randomly selected from T to form a new sample set T^*,(x_1^*, y_1^*), (x_2^*, y_2^*), … , (x_n^*, y_n^*), as the sample at the root node of the decision tree.

(3) Randomly select m attributes from M attributes as candidate attributes of a single tree, and satisfy the condition $m << M$. Then a certain strategy (information gain is used here) from these m attributes is used to select 1 attribute as the split attribute of the node.

① Information entropy:

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \quad (3.5)$$

$$p_i = \frac{|C_i|}{|D|} \quad (3.6)$$

Where X represents a discrete random variable, p_i represents the probability of each random variable in the entire event, $|D|$ represents the sample size, and $|C_i|$ represents the number of samples belonging to the C_i category.

② Information gain:

$$g(Y|X) = H(Y) - H(Y|X) \quad (3.7)$$

Where $\mathrm{H(Y)}$ is the empirical entropy of sample category Y, and $\mathrm{H(Y|X)}$ is the empirical conditional entropy.

(4) In the process of decision tree formation, each node must be split according to step 3 (if the next attribute selected by the node is the attribute that was just used when the parent node was split, then the node has reached the leaf node, There is no need to continue to split). Until it can no longer be divided. Note that there is no pruning during the entire decision tree formation process.

(5) Follow steps 3 to 4 to establish t decision trees to form a random forest.

(6) For the input vector x_i, each tree outputs the classification result for voting.

(7) Counting the results of voting, the category with the highest number of votes is the category label of x_i.

The intensity of tree classification is defined as:

$$\mathrm{s} = E_{X,Y} mr(X, Y) \quad (3.8)$$

The upper bound of the generalization error of the random forest tree is given by:

$$PE^* \le \bar{\rho}(1 - s^2) / s^2 \quad (3.9)$$

Among them, $\bar{\rho}$ measures the average correlation between various variables.

It can be seen from the above formula that the error of the random forest tree algorithm depends on the correlation between each tree.

For the random forest model, we build a model based on the molecular descriptors and biological activity after pre-processing.

3.5 Random forest model solution and analysis

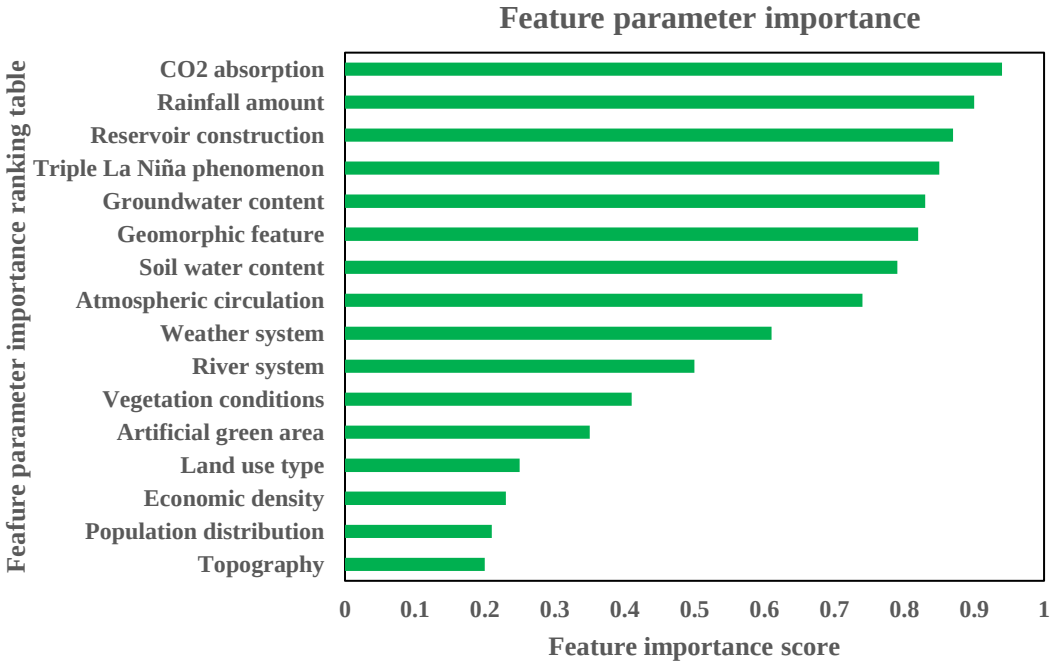


Fig. 3.2 Ranking chart of feature importance

The importance of the variables to the heat and drought disasters in southern China was solved by Python programming and sorted to obtain Fig. 3.2. Here, the forest area representation is used to show the importance ranking of the 16 variable characteristics affecting the heat and drought disasters in southern China. Combined with the actual problems, the model selects the top 9 most significant variables.

3.6 Prediction model solution and analysis

By establishing the grey correlation model, taking the heat and drought disasters in southern China as the reference columns, the influence of the variable characteristic indexes including the Triple La Niña phenomenon on the forest land area in southern China was analyzed. The analysis finds that the first nine variables are more important to the heat and drought disasters in southern China, and the importance is in the forefront. Table 3.1 shows the changes in the order of importance of heat and drought disasters in southern China.

Table 3.1 Variation scale of importance ranking of thermal and drought disasters in southern China

The order	A(10^4t)	B(mm/year)	C(m^3)	D(day/year)
Variable	CO_2 absorption	Rainfall amount	Reservoir construction	Triple La Niña phenomenon
The order	E(10^4m^3)	F	G(%)	H(hm^2)

Variable	Groundwater content	Geomorphic feature	Soil water content	Atmospheric circulation
The order	I	J	K	L
Variable	Weather system	River system	Vegetation conditions	Artificial green area
The order	M	N(10^{12}\$)	O(M)	P
Variable	Land use type	Economic density	Population distribution	Topography

Then, the random forest model was used to analyze the ranking of variable characteristic parameters. The 16 molecular descriptors with the most significant influence on biological activity were visualized by the ranking table. It was found that the first 16 variables were consistent with the importance variables screened by the grey correlation model, which also verified the rationality and reliability of the first 9 molecular descriptors selected by us. Thus laid the foundation for the analysis of the following issues.

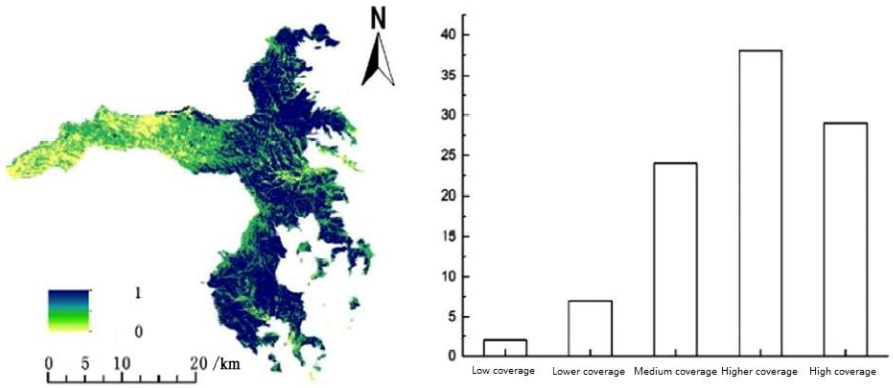


Fig. 3.3 Spatial distribution characteristics and area proportion of average vegetation coverage in South China Forest Farm from 1980 to 2021

By querying relevant data, we have produced a map of the spatial distribution characteristics and area proportion of the average vegetation coverage of forest farms in southern China from 1980 to 2021. According to Fig. 3.3, the average vegetation coverage of forest farms in southern China from 1980 to 2021 is high, and the low coverage only accounts for 2% of the total area. The spatial characteristics of vegetation coverage in the study area are obvious, which is high in the east and low in the west. This is because the western part of the region soil desertification is serious, afforestation task is arduous and slow progress, resulting in low vegetation coverage in the western region.

Use Box-Behnken Design (BBD) method to analyze the acquired experimental data, and take the forest area X as the response value, as the independent variables B, C, G, H, J, L, O, P (in the symbol Corresponding indicators have been explained in the

description) The response surface quadratic polynomial for forest land area prediction is established, as shown in (formula).

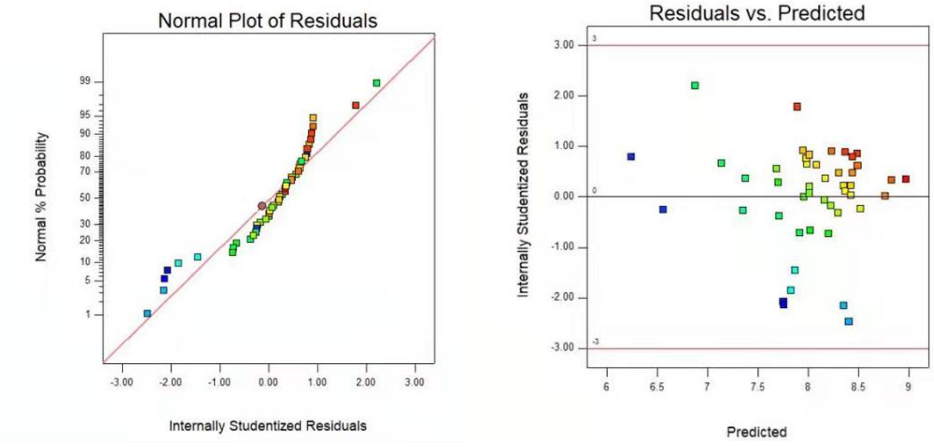
$$X = -42296.42 + 5880.43\ \text{B-}2852.49\text{D} + 4058.77\text{G} - 8.19\text{H} + 4046.14\text{J} + 1062.16\ \text{L} - 52.73\text{O-}2850.71\text{P}$$

In the response surface analysis process, the F value is used to detect the significance of the statistical results, and the p value is used to detect the significance of the regression coefficient. The smaller the p value, the more significant the result. Table 3.2 has shown part of the response surface analysis results, and the specific analysis results are shown in the attachment. From Table 3.2, the F value of the model is 3.19, $p<0.016$, indicating that the model has significant adaptability. The nonlinear relationship between the factors in the regression equation and the response value is significant, that is, the model has a high reliability. Can be used to predict X.

Table 3.2 Response surface Analysis Table (part)

parameter	Sum of squares	The mean square	The F value	The P values
model	0.135	0.070	3.19	0.0154(noticeable)
B	0.014	0.014	0.71	0.3952
C	0.052	0.053	2.53	0.1324
G	0.027	0.32	13.70	0.0021
O	0.021	0.21	8.61	0.0125
residual	0.314	0.024		
Net error	0.14	0.030		
Total error	1.40	-		

In general, when analyzing the data, the closer the probability of the residual normal distribution is to a straight line, the residual is predicted by the equation. The more confusing the value correspondence, the higher the credibility of the model. Through analysis, the corresponding relationship obtained by this model is shown in Fig. 3.4.



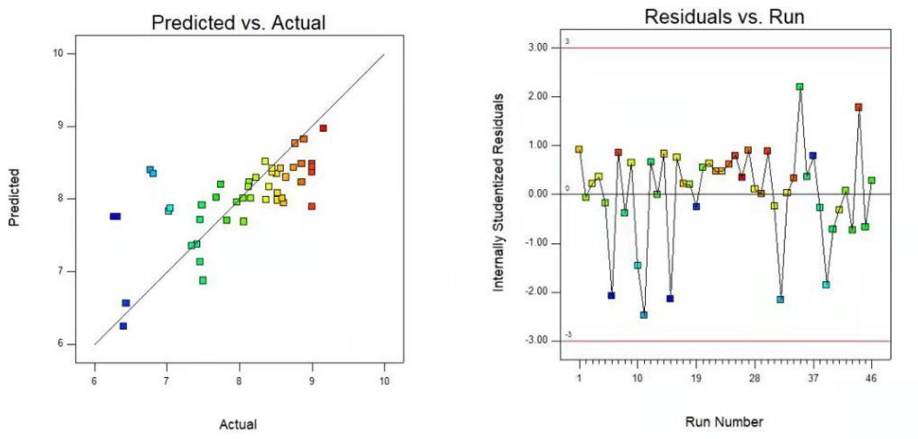


Fig. 3.4 Residual positive distribution and Residual vs predict

From Fig. 3.4, it can be seen that the residual, normal, and probability distribution curves obtained by this model are basically close to the straight line, and the distribution is reasonable. The correlation between the residuals and the equation prediction has a high degree of dispersion. It is calculated by the response surface quadratic polynomial. The predicted value and the true value are basically close to the same straight line, indicating that the model obtained by this method has a certain degree of accuracy.

Next, draw a two-factor relationship diagram (partially) to analyze each factor, as shown in Fig. 3.5.

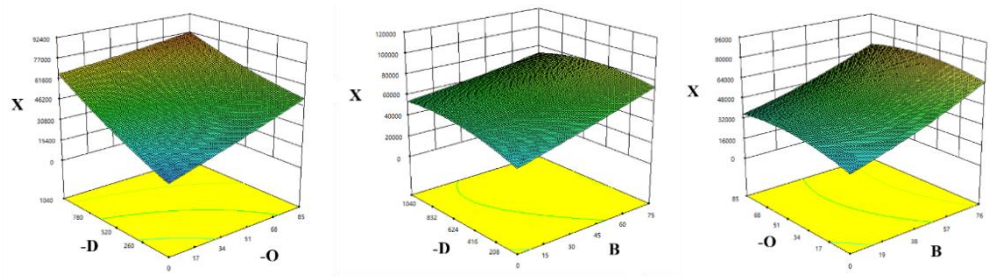


Fig. 3.5 Factor -D,-O diagram; Factor -D,B diagram; Factor -O,B diagram

From Fig. 3.5, it can be seen that the woodland area X changes significantly with B-annual rainfall, D-duration of Triple La Niña phenomenon, and O-population distribution, B is positively correlated with woodland area, D, and O are negatively correlated with woodland area which is in line with the conclusion drawn from problem one; in general, the correlation degree of the above three factors on woodland area can still be expressed as: B>D>O.

The above process argues the reasonableness and scientificity of the model, and the influencing factors are analyzed in detail, and the model is next used to forecast the forest area. Response Surface Methodology (Response Surface Methodology, RSM) obtains the test result data by designing a reasonable test method and through a reasonable operation, and uses a multiple quadratic linear regression method to establish the functional relationship between the independent variable and the response value. Analyze the change trend of the response value. This method is more common

in solving multivariate problems.

This modeling uses the world's top experimental design software Design-Expert, and the Box-Behnken Design (BBD) experimental design method. In the design process, the response surface refers to the functional relationship between the response variable (dependent variable) y and a set of input variables $x_1, x_2, ..., x_n$, which satisfies $y = f(x_1, x_2, ..., x_n)$.

Table 3.3 Table of Forestland Area Scale required by Each Province (Response surface Model)

Provinces	Beijing	Tianjin	Hebei	Shanxi
Construction area(hm^2)	143.517	369.631	1229.551	3079.986
Liaoning	Shanghai	Jiangsu	Zhejiang	Anhui
760.648	442.690	1692.240	469.885	602.045
Shandong	Henan	Guangdong	Guizhou	Shaanxi
2958.204	1103.026	391.306	20.635	666.665
Qinghai	Ningxia	Xinjiang		
20.317	634.182	344.083		

3.7 Summary of question 2

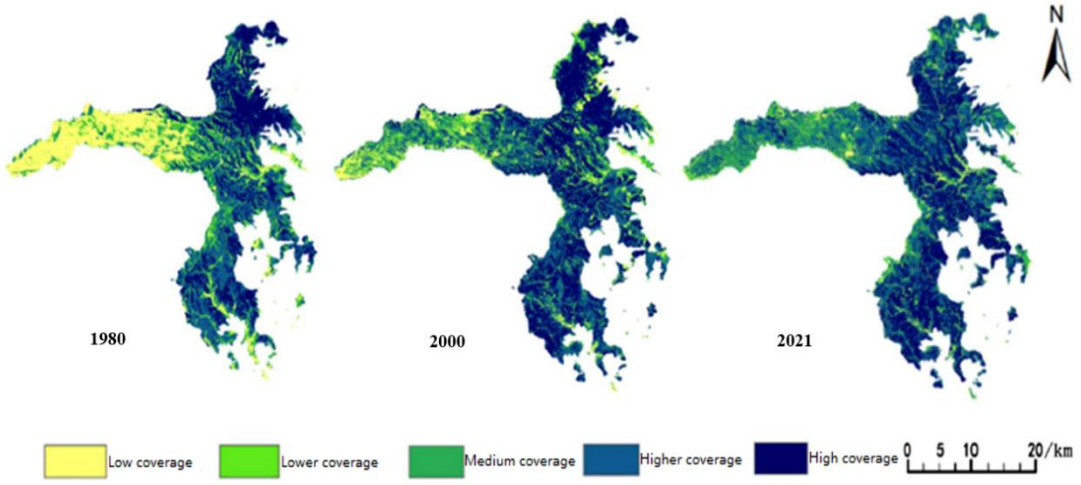


Fig. 3.6 Spatial distribution of vegetation coverage in South China in 1980,2000 and 2021

Table 3.4 Proportion of vegetation coverage in Southern China Mechanical Forest Farm in three periods

Vegetation coverage level	Area than		
	In 1980	In 2000	In 2021
Low coverage	14	9	7

Lower coverage	7	8	4
Medium coverage	16	23	16
Higher coverage	29	30	32
High coverage	34	30	28

The spatial characteristics of vegetation cover in 1980, 2000 and 2021 in southern China's forests are shown in Fig. 3.6. From Fig. 3.6, it can be seen that the vegetation cover in southern China improved significantly over time from 1980 to 2021, with low and lower coverage gradually developing into medium and higher coverage. It can be clearly seen that the vegetation coverage of medium coverage class in the forestry field changes significantly and in a decreasing trend over time, the overall change in the percentage of low and higher coverage area is relatively smooth, the percentage of low coverage area, medium coverage area and high coverage area decreases as time proceeds, and there is no stable trend in the change of other classes, Triple La Niña phenomenon, CO_2 absorption, and rainfall reduction significantly slow down the change of medium vegetation coverage of the area with high coverage, this winter Triple La Niña in southern China may start in December 2022 and last until February 2023.

4. Problem 3 model establishment and solution

4.1 Analysis of problem 3

For question three, the question requires the selection of suitable indicators to develop an evaluation model of various disaster losses caused by flooding under Triple La Niña, in order to quantitatively evaluate the impact of Triple La Niña and propose targeted response strategies. Before building the model, the data obtained from the relevant literature and national statistical office research were partially discarded by pre-processing the data and the number of variables was reduced by multidimensional. In this problem, the variable with the greatest and most representative impact of Triple La Niña is selected, and the processed data is analysed and predicted through a prediction model, and the predicted results are compared with the actual data to quantify the assessment.

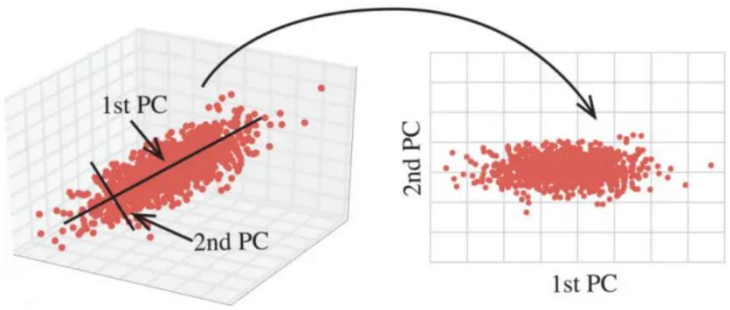


Fig. 4.1 Dimensionality reduction diagram.

For the forecasting model, two main forecasting methods were considered in the early stage, namely (i) time series analysis method: using moving average method to forecast the change of industry growth rate in the next four quarters; and (ii) neural network. Before making the forecast, the two methods were compared for their suitability for this problem and we finally chose the first method for forecasting and evaluation.

4.2 Establishment of evaluation model

4.2.1 Time series model is established

Through our research, we have collected data on the income of various industries in various countries for several years, and based on various considerations such as the countries involved in the Triple La Niña induced floods and the volume of data, we have finally taken Australia as an example.

Fig. 4.2 Flood affected areas in Australia.

By arranging the percentage increase in output value of each industry in chronological order, a time series is formed and the trend and pattern of future changes are deduced from the past changes in the formed time series, which is the time series forecasting method. This method is in fact a regression model, which considers that the change in growth rate has a continuous nature, so it can be used to predict the pattern and trend of change through statistical analysis of past data. This method is simple to operate, easy to grasp, fast to calculate, and can be used for preliminary analysis and general analysis of change trends, but this method also has certain limitations, for example, this method can only be used for short time forecasting, and cannot reflect the internal rules and links of development, etc. The prediction time in this problem is short compared to the data already collected, so the time series approach can be applied.

To predict a time series, we first need to discover the patterns of its changes. The most basic way to predict the future trend of a time series is to find the cycles of time series changes by mining the evolutionary features of the time series data through

manual experience. Specifically looking at a time series, when the series is cyclical, the cyclical features of the time series are extracted for prediction.

Through researching the literature, we found that a large number of forecasts are made using the moving average method, which has the advantages of simplicity and accuracy. The average of the values of the most recent N periods of the series is used as the forecast for each future period. Generally N is taken to be in the range 5 $\leqslant$ N $\leqslant$ 200. The growth rate data for each sector did not vary much, but the underlying trend of the data was difficult to capture, so the value of N should be taken as large as possible, and N=10 was finally determined.

The moving average is chosen for this analysis, and its basic principle: let the observed series be $y_1,y_2,\ldots,y_T$, take the number of terms of the moving average N < T, when the basic trend of the forecast target is fluctuating up and down at a certain level, the forecasting model can be built by one moving average method as follows.

$$\hat{y}_{t+1} = M_t^{(1)} = \frac{1}{N}\left(y_t + \ldots + y_{t-N+1}\right), t = N, N+1, \ldots, T \tag{4.1}$$

The standard error of prediction for this forecasting method is

$$S = \sqrt{\frac{\sum_{t=N+1}^{T}\left(\hat{y}_t - y_t\right)^2}{T-N}} \tag{4.2}$$

4.2.2 Principal component analysis model establishment

There are various types of disaster losses caused by flooding under the effect of Triple La Niña events, such as agricultural losses, housing losses, etc. From the data obtained from our research, we selected typical indicators to quantify and evaluate the disaster losses. Due to the large number of quantitative indicators and the large amount of data, we chose a principal component analysis model to evaluate them. The five main sectors selected as indicators are agriculture, forestry and fishing, manufacturing, construction, wholesale trade and retail trade volume growth.

Next, the model building process is introduced.

Principal component analysis evaluation index Here we select the five indexes selected in question three, and use $x_1,x_2,\ldots,x_5$ to denote the number of years i respectively. The value of $i=1,2,3,\ldots,n$ in the i year is denoted as $x_1,x_2,\ldots,x_5$ respectively, and the matrix $[a_{i1},a_{i2},\ldots,a_{i5}]$ is constructed.

The evaluation steps based on principal component analysis are as follows:

(1) Standardize the raw data. Convert each indicator value a_{ij} into a standardized indicator a_{ij}, there are

$$a_{ij} = \frac{a_{ij} - \mu_j}{s_j}, i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, 5 \tag{4.3}$$

In the formula, $\mu_j = \frac{1}{n}\sum_{i=1}^{n} a_{ij}; s_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(a_{ij} - \mu_j\right)^2}$, $j = 1, 2, \ldots, 5$, μ_j ,s_j is the sample mean and sample standard deviation of the j index. Correspondingly, said:

$$x_j = \frac{x_j - \mu_j}{s_j}, i = 1,2,\ldots,5 \quad (4.4)$$

is a standardized variable.

(2) Calculate the correlation matrix R, whose correlation coefficient matrix $R = \left(r_{ij}\right)_{5*5}$ satisfies:

$$r_{ij} = \frac{\sum_{k=1}^{n} a_{ki} \cdot a_{kj}}{n-1}, i, j = 1,2,\ldots,5 \quad (4.5)$$

Where: r_{ij}=1; $r_{ij} = r_{ij}$; r_{ij} is the correlation coefficient between the *i* index and the *j* index.

(3) Calculate eigenvalues and eigenvectors. Calculate the eigenvalue $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_5 \geq 0$ of the correlation coefficient matrix R and the corresponding standardized eigenvector $u_1, u_2, \ldots, u_5$, where $u_j = \left[u_{1j}, u_{2j}, \ldots, u_{5j}\right]^T$ is composed of 5 new index variables by the eigenvectors:

$$\begin{aligned} y_1 &= u_{11}x_1 + u_{21}x_2 + \ldots + u_{51}x_5, \\ y_2 &= u_{12}x_1 + u_{22}x_2 + \ldots + u_{52}x_5, \\ &\cdots \\ y_5 &= u_{15}x_1 + u_{25}x_2 + \ldots + u_{55}x_5, \end{aligned} \quad (4.6)$$

In the formula, y_1 is the first principal component; y_2 is the second principal component; ...; y_5 is the fifth principal component.

(4) Select $p(p \leq 5)$ principal components and calculate the comprehensive evaluation value.

Calculate the information contribution rate and cumulative contribution rate of the characteristic value $\lambda_j (j = 1,2,\ldots,5)$. say:

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{5} \lambda_k}, j = 1,2,\ldots 5 \quad (4.7)$$

The information contribution rate of the main component y_j;

$$\alpha_p = \frac{\sum_{k=1}^{p} \lambda_k}{\sum_{k=1}^{5} \lambda_k} \quad (4.8)$$

It is the cumulative contribution rate of the main component $y_1, y_2, \ldots, y_p$.

4.3 Time series model solution and analysis

After data pre-processing, the data completeness of the quantitative variables analysed is more complete, up to 90% or more, and the high data completeness increases the credibility of the model solution results.

Using this method, the change in the growth rate of the selected industry for the next four quarters was obtained. The object of this forecast is the industry growth rate, whose value reflects the development of the industry, the larger the value, the better the

industry is developing in that quarter.

From the analysis of the forecast curve, it is found that the forecast curve maintains the growth trend of the original curve data with an upward pattern, but the forecast results are not entirely ideal and there are certain problems, such as the forecast results cannot reflect the oscillating nature of the original curve and the forecast results have certain bias due to the large oscillation of the original data.

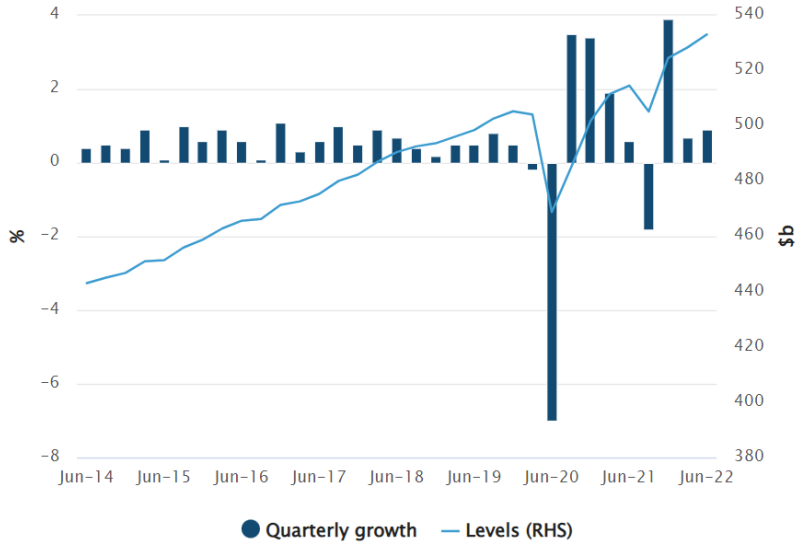


Fig. 4.3 Gross domestic product, chain volume measures, seasonally adjusted

Common sense dictates that the five sectors of agriculture, forestry and fishing, manufacturing, construction, wholesale trade and retail trade were more significantly affected by the disaster than the education and training, accommodation and catering sectors. In order to quantify the fluctuating trends in the data, we have collected the changes in the country's GNP from 1960 to the present, which shows a more pronounced trend. Looking at the curve reveals that its GDP should be on an upward trend for the four quarters of 2022, and comparing it with the actual situation reveals a large gap, by which we can determine that Australia suffered more from the flooding disaster caused by the Triple La Niña. Comparing the time series forecast curve with the GNP forecast curve, it is found that the trend of the time series forecast is much flatter than it is, mainly because the time series method does not take into account the variation in the effects between different parameters and does not closely integrate their intrinsic patterns and linkages, hence the bias in the forecast results.

4.4 Summary of question 3

Combining the data and forecast curves we can observe that the more obviously affected sectors, particularly in the spring agricultural harvest, agriculture, forestry and fishing did not show significant growth, and manufacturing and other sectors also showed a not insignificant fall due to the suspension of work in the floods. To this end we believe that the government should provide assistance to primary producers and rural landowners, increase the amount of credit available to small and medium-sized

enterprises to help revive them and, where appropriate, provide funding for housing and accommodation.

Statistically, Australia has experienced three major floods in the first three quarters of the year and several states are still experiencing high levels of flooding events this season. To this end, we recommend that the government prepare a dedicated disaster preparedness fund to curb the devastating effects of natural disasters, including but not limited to direct investment in infrastructure, such as levees, floodways, seawalls, artificial wetlands, etc., with the aim of reducing systemic risk to improve resilience to future disasters. At the same time investment in flood risk management should be increased, including but not limited to environmental restoration and clean-up.

Finally, we find out that floods occur mainly in the north and south of Australia and that the northern river regions are vulnerable to severe weather, so we propose to carry out a southern and northern river restoration, using their climate, catchment areas and hydrological drivers as parameters to find internal links between them and floods through impact analysis.

5. Problem 4 solution

5.1 Analysis of problem 4

Question 4 requires a report to the relevant management on the triple La Nina incident. This problem can be combined with the content of the previous three problems to explain our understanding of the Triple La Niña event to the National Weather Service and give corresponding countermeasures.

5.2 Report to the Met Office

Floods and heatwaves are currently causing climate disruptions across the globe. As the global climate warms, the frequency of extreme weather events is changing, showing an increasing trend. In fact, humans are adapting to extreme weather events, with hot summer heat, violent hurricanes, biting cold and torrential floods becoming "regular" events, and good weather is now considered a luxury. According to the Hong Kong Observatory, the frequency of extreme weather events caused by global warming is expected to increase worldwide since 2013.

After many days of struggle by the team, a statistical analysis of the Triple La Niña impact areas was carried out to clarify the main impact areas and Chengdu of the Triple La Niña, and the corresponding climate change characteristics were obtained. Triple La Niña events mainly affect regions such as Australia, the United States, Canada, China, northeastern Brazil, India and southern Africa, and the South American coast, bringing natural disasters such as floods, droughts and changes in precipitation, which in turn affect local economic production.

The spatial distribution of ocean temperatures at different times has been mapped using ocean temperature data, which shows a global warming trend over most of the

globe, with the greatest increase in temperature in the Indian Ocean Southeast Asia, a significantly stronger temperature change in the northern hemisphere and equatorial regions than in the southern hemisphere, and an unstable change in temperature in the equatorial Pacific, with areas of low temperature, which is closely linked to a Triple La Niña event, which will most likely occur in the future and last until December 2022 to February 2023.

Drawing on our drought and the handling of floods in Australia, we ultimately concluded that the government should make good observations and provide early warning of possible disasters; set up funds in advance to provide assistance to primary producers and rural landowners in the event of a disaster, lines of credit to small and medium-sized enterprises and, where appropriate, funding for housing and accommodation. It is also recommended that the government should prepare a dedicated disaster preparedness fund to curb the devastating effects of natural disasters, including but not limited to direct investment in infrastructure such as levees, floodways, seawalls, artificial wetlands, etc., with the aim of reducing systemic risk and increasing resilience to future disasters. Investments in risk management for droughts and floods should also be increased, including but not limited to environmental restoration and clean-up.

6. Strengths and Weakness

6.1 Strengths of the model

Since quantitative data on global temperature, drought, and floods are extremely difficult to obtain, we start from the source and use the impact of Triple La Niña on global sea surface temperature, forest area in southern China, and economic data of Australian industries as evaluation models to make the models more realistic and reliable.

For predicting the impact of Triple La Niña for different disasters, we chose different prediction models based on historical trends to make the model results with high accuracy.

The successive approximation expression of wavelet analysis is determined by wavelet analysis theory, which can avoid the blindness of structural design such as BP neural network. We chose the response surface analysis model to analyze the three measures. This method enables the simultaneous analysis of multiple measures and presents the results in the form of 3D plots, which are intuitive and specific. Neural network has strong adaptability, fault tolerance and self-organizing ability.

6.2 Weakness of the model

(1) Due to the limited amount of calculation, the selected indicators in this paper are limited. In the next step, more indicators can be considered for evaluation and reasonable classification of indicators can be carried out to increase the reliability of the model.

(2) Since the query index data is messy, we chose the index with relatively high integrity rate for analysis, which needs to be further improved in the subsequent optimization.

7. Conclusion

Through the establishment of wavelet analysis model, it is concluded that there are multiple oscillation periods in the ocean temperature time series, and the overall trend is upward. The maximum main period of the time series is 6 months, the second main period is 112 months, and the oscillation is the strongest under the maximum main period. In the first main cycle trend diagram of the six-month time scale, the average cycle is shortened from 180 months in the early stage to 80 months in the later stage, indicating that the ocean temperature changes from a relatively stable state to an unstable state. This phenomenon may be related to human factors such as greenhouse gas emissions. In the second main cycle of the 112-month time scale, the average period ranges from 250 months to 440 months, and the contribution rate is small. This phenomenon may be related to natural factors such as uncertain periodic activity, Pacific interdecadal oscillation, La Niña phenomenon, and the period range of the second main cycle is consistent with the Pacific interdecadal oscillation period obtained from the survey, which further verifies the correctness of the speculation.

Using the ocean temperature data, the spatial distribution map of ocean temperature at different times is drawn, and it is concluded that the temperature in most parts of the world is on the rise. Among them, the temperature rise in the southeast Asia of the Indian Ocean is the largest. The temperature change in the northern hemisphere and the equatorial region is significantly stronger than that in the southern hemisphere. The temperature in the equatorial Pacific region shows an unstable change and there is a low temperature region, which is closely related to the Triple La Niña event. In the future, the Triple La Niña event is very likely to occur and continue until December 2022 to February 2023.

From 1980 to 2021, the vegetation cover in southern China has improved significantly over time, with low and low cover gradually developing into medium and high cover. According to the gray correlation model and the Box-Behnken Design (BBD) method, it was concluded that the vegetation cover of the medium cover class in the forestry field changed significantly and showed a decreasing trend over time, and the overall change of the ratio of low and high cover area was relatively stable, and the ratio of low cover area, medium cover area, and high cover area decreased over time. While there is no stable trend in the changes of other classes, the Triple La Niña phenomenon, CO_2 absorption, and reduced rainfall obviously slow down the changes of medium vegetation in high-coverage areas. The model established by forest area predicts that the Triple La Niña phenomenon in southern China this winter may start from December 2022 and last until February 2023.

Common sense dictates that the five sectors of agriculture, forestry and fishing, manufacturing, construction, wholesale trade and retail trade were more significantly affected by the disaster than the education and training, accommodation and catering sectors. In order to quantify the fluctuating trends in the data, we have collected the

changes in the country's GNP from 1960 to the present, which shows a more pronounced trend. Looking at the curve reveals that its GDP should be on an upward trend for the four quarters of 2022, and comparing it with the actual situation reveals a large gap, by which we can determine that Australia suffered more from the flooding disaster caused by the Triple La Niña. Comparing the time series forecast curve with the GNP forecast curve, it is found that the trend of the time series forecast is much flatter than it is, mainly because the time series method does not take into account the variation in the effects between different parameters and does not closely integrate their intrinsic patterns and linkages, hence the bias in the forecast results.

References

[1] Xu Y , Ramanathan V , Victor D G . Global warming will happen faster than we think[J]. Nature, 2018, 564(7734):30-32.

[2] Medhaug I , Tolpe M B S , Fischer E M , et al. Reconciling controversies about the 'global warming hiatus'[J]. Nature, 2017, 545(7652):41.

[3] Azmoodehfar M H , Azarmsa S A . Assessment the Effect of ENSO on Weather Temperature Changes Using Fuzzy Analysis (Case Study: Chabahar)[J]. Apcbee Procedia, 2013, 5:508-513.

[4] Chen X J , Zhao X H , Chen Y . Influence of El Nino / La Nina on the western winter-spring cohort of neon flying squid (Ommastrephes bartramii) in the northwestern Pacific Ocean[J]. Ices Journal of Marine Science, 2007, 64(6):1152-1160.

[5] Chen X , Guo Y P , Tan Z M , et al. Influence of different types of ENSO events on the tropical cyclone rainfall over the western North Pacific[J]. Climate Dynamics, 2022:1-14.

[6] Jones N . Rare 'triple' La Nia climate event looks likely — what does the future hold[J]. Nature, 2022, 607.

[7] Zheng F , Liu J P , Fang X H , et al. The Predictability of Ocean Environments that Contributed to the 2020/21 Extreme Cold Events in China: 2020/21 La Nina and 2020 Arctic Sea Ice Loss.

[8] Xiao-Chen Yuan et al. Risk analysis for drought hazard in China: a case study in Huaibei Plain[J]. Natural Hazards, 2013, 67(2) : 879-900.

Appendix

1. Data of Southern China from 1962 to 2021

Year	Sand and dust weather times/d	Average number of triple La Niña influenced /d	Total forest reserves/ 10^4m^3	Average annual precipitation/mm	Forest cover rate/%	CO_2 absorption /10^4t
2021	3	68	1030.1	486	82.6	75.29
2020	2	59	1032.8	479	82	74.71
2019	3	65	1012.3	475	81.2	70.06
2018	2	56	922.2	474	79.1	69.87
2017	3	52	810.4	429	78.3	66.24
2016	5	53	789.2	548	77.4	65.32
2015	6	55	783.2	432	75.3	63.34
2014	5	60	670.3	328	74.4	56.3
2013	4	56	650.3	430	73.2	53.2
2012	1	54	660.2	413	71.3	51.3
2011	4	57	639.2	372	69.4	50.23
2010	8	55	632.2	442	68.3	49.89
2009	3	56	622.3	302	67.2	50.34
2008	6	57	563.2	402	65.3	45.34
2007	5	58	574.2	322	64.2	44.32
2006	22	59	550.5	390	62.3	43.23
2005	4	58	530.3	382	60.34	41.34
2004	7	60	539.4	454.2	59.34	40.23
2003	1	62	489.3	452	58.3	40.03
2002	29	63	487.5	339	57.3	39.23
2001	24	64	468.2	342	56.3	35.23
2000	22	63	452.3	352	55.24	34.98
1999	5	62	443.2	428	54.23	33.22
1998	2	61	439.2	528	52.21	32.23
1997	0	62	421.3	402	50.5	31.39
1996	8	63	421.2	456	48.34	32.12
1995	8	61	411.3	455	47.45	29.23
1994	6	63	401.1	464	45.32	28.34
1993	12	62	401.2	472	44.45	27.34
1992	9	63	398.1	480	43.24	27.09
1991	10	65	392.2	491	42.1	26.99
1990	14	66	350.3	482	40.3	25.32
1989	2	69	338.1	298	39.45	24.34
1988	24	75	320.5	332	38.45	24.01
1987	11	73	300.3	423	37.54	23.34

1986	15	74	280.7	478	36.55	23.03
1985	15	76	240.8	452	34.84	21.13
1984	19	75	170.3	301	34.53	21.01
1983	22	81	160.7	432	33.95	20.11
1982	23	84	150.5	372	31.95	19.23
1981	27	82	129.9	350	30.45	18.88
1980	32	84	100.4	302	26.68	17.23
1979	34	82	78.6	505	24.43	15.34
1978	36	85	56.3	485	20.34	14.23
1977	19	69	53.2	391	18.32	14.01
1976	44	89	55.3	401	15.43	13.2
1975	29	79	56.3	332	13.99	13.11
1974	35	90	55.3	481	13.89	12.23
1973	25	81	54.3	498	13.59	12.01
1972	20	82	50.2	328	13.45	11.23
1971	22	84	49.4	348	13.3	11.11
1970	17	72	45.5	423	13.1	11.02
1969	24	83	45.5	502	12.8	10.87
1968	38	80	42.4	356	12.5	10.78
1967	26	85	41.3	405	12.1	10.55
1966	43	93	40.3	370	11.9	10.45
1965	44	90	39.4	382	11.8	9.99
1964	8	73	38.7	452	11.4	9.89
1963	21	82	35.7	408	11.3	9.45
1962	20	83	34.5	386	11.4	8.89

2. Forest area in five provinces of China from 2004 to 2019($10^4 hm^2$)

Years	Hebei	Shanxi	Jiangsu	Shandong	Henan
2019	502.69	321.09	155.99	266.51	403.18
2018	502.69	321.09	155.99	266.51	403.18
2017	502.69	321.09	155.99	266.51	403.18
2016	502.69	321.09	155.99	266.51	403.18
2015	502.69	321.09	155.99	266.51	403.18
2014	502.69	321.09	155.99	266.51	403.18
2013	439.33	282.41	162.1	254.6	359.07
2012	439.33	282.41	162.1	254.6	359.07
2011	439.33	282.41	162.1	254.6	359.07
2010	439.33	282.41	162.1	254.6	359.07
2009	439.33	282.41	162.1	254.6	359.07
2008	418.33	221.11	107.51	254.6	336.59
2007	418.33	221.11	107.51	254.6	336.59
2006	418.33	221.11	107.51	254.6	336.59
2005	418.33	221.11	107.51	254.6	336.59
2004	418.33	221.11	107.51	254.6	336.59

3. The forest area of each province in Indonesia from 2016 to 2018

Provinces	2016	2017	2018	2016	2017	2018	2016	2017	2018
ACEH	3352.8	3352.8	3349.7	2294.5	2294.5	2297.6	5647.3	5647.3	5647.3
SUMATERA UTARA	3055.8	3055.8	3055.8	4046.2	4046.2	4046.2	7102.0	7102.0	7102.0
SUMATERA BARAT	2342.9	2342.9	2342.9	1841.0	1841.0	1841.0	4183.9	4183.9	4183.9
RIAU	5407.0	5407.0	5407.0	3475.8	3475.8	3475.8	8882.8	8882.8	8882.8
JAMBI	2098.5	2098.5	2098.5	2733.8	2733.8	2733.8	4832.3	4832.3	4832.3
SUMATERA SELATAN	3408.8	3408.8	3407.7	5218.1	5218.1	5219.2	8626.9	8626.9	8626.9
BENGKULU	924.6	924.6	924.6	1078.3	1078.3	1078.3	2002.9	2002.9	2002.9
LAMPUNG	1004.7	1004.7	1004.7	2430.6	2430.6	2430.6	3435.4	3435.4	3435.4
KEP. BANGKA BELITUNG	654.6	643.6	643.6	1005.2	1016.1	1016.1	1659.7	1659.7	1659.7
KEP. RIAU	382.1	382.1	381.8	434.9	434.9	435.2	817.0	817.0	817.0
DKI JAKARTA	0.5	0.5	0.5	64.9	64.9	64.9	65.3	65.3	65.3
JAWA BARAT	816.6	816.6	816.6	2882.0	2882.0	2882.0	3698.6	3698.6	3698.6
JAWA TENGAH	647.1	647.1	647.1	2809.4	2809.4	2809.4	3456.6	3456.6	3456.6
DI YOGYAKARTA	16.8	16.8	16.8	302.6	302.6	302.6	3194.0	319.4	319.4
JAWA TIMUR	1357.6	1357.6	1357.6	3480.0	3480.0	3480.0	4837.7	4837.7	4837.7
BANTEN	201.8	201.8	201.8	737.4	737.4	737.4	939.2	939.2	939.2
BALI	127.3	127.3	127.3	439.6	439.6	439.6	566.9	566.9	566.9
NUSA TENGGA	1035.8	1035.8	1035.8	944.3	944.3	944.3	1980.2	1980.2	1980.2

RA BARAT									
NUSA TENGGARA TIMUR	1485.9	1485.9	1485.9	3236.6	3236.6	3236.6	4722.5	4722.5	4722.5
KALIMANTAN BARAT	8198.7	8198.7	8198.7	6374.1	6374.1	6374.1	14572.8	14572.8	14572.8
KALIMANTAN TENGAH	12697.2	12697.2	12697.2	2569.0	2569.0	2569.0	15266.2	15266.2	15266.2
KALIMANTAN SELATAN	1780.0	1780.0	1780.0	1934.0	1934.0	1934.0	3713.9	3713.9	3713.9
KALIMANTAN TIMUR	13855.8	13833.1	13833.1	5649.0	5671.7	5671.7	19504.8	19504.8	19504.8
SULAWESI UTARA	694.9	694.9	694.9	744.6	744.6	744.6	1439.5	1439.5	1439.5
SULAWESI TENGAH	3934.6	3934.6	3934.6	2100.1	2100.1	2100.1	6034.7	6034.7	6034.7
SULAWESI SELATAN	2119.0	2119.0	2119.0	2379.4	2379.4	2379.4	4498.4	4498.4	4498.4
SULAWESI TENGGARA	2326.4	2326.4	2326.4	1285.2	1285.2	1285.2	3611.6	3611.6	3611.6
GORONTALO	824.7	824.7	824.7	373.8	373.8	373.8	1198.5	1198.5	1198.5
SULAWESI BARAT	1092.4	1092.4	1092.4	587.9	587.9	587.9	1680.2	1680.2	1680.2
MALUKU	3910.4	3910.4	3910.4	711.7	711.7	711.7	4622.1	4622.1	4622.1
MALUKU UTARA	2515.2	2515.2	2515.2	615.4	615.4	615.4	3130.6	3130.6	3130.6
PAPUA BARAT	8784.8	8784.8	8784.8	840.1	840.1	840.1	9624.9	9624.9	9624.9