RESEARCH

RSST-ARGM: A Data-Driven Approach to Long-term Sea Surface Temperature Prediction

Linqian Zhu[1*], Qi Liu[2], Xiaodong Liu[3] and Yonghong Zhang[4]

[*]Correspondence: zderrick28@163.com
[1]School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China
Full list of author information is available at the end of the article

Abstract

For the purpose of exploring the long-term variation of regional SST, this paper studies the historical SST in local sea areas and the emission pattern of greenhouse gases and proposes a gray model of regional SST based on atmospheric reflection which can be used to predict SST variation in a long time span. By studying the grey systematic relationship between historical SST data, the model obtains the development law of temperature change, and further introduces different future greenhouse gas emission scenarios as the index coefficient to determine the corresponding changing results of seawater temperature in the next 50 years. Taking the North Atlantic Ocean as an example, the cosine similarity test method is used to verify the model proposed in this paper, and its accuracy is as high as 0.99984. The model predicts that the local SST could reach a maximum of 15.3℃ by 2070. This model is easy to calculate, with advantages of the high accuracy and good robustness.

Keywords: Long-term prediection; Regional SST; Temperature variation; Gray model; Atmospheric reflection

1 Introduction

The ocean is one of the most critical areas to maintain the earth's ecology, which is rich in natural resources and economic and social values. The trend of ocean temperature change in the future will have great importance in environmental protection as well as industrial development. The characteristics of regional SST change accord with the gray system, combining theories of the gray system and RCPs index. The prediction model of regional SST under atmospheric radiation RSST-ARGM is built. Based on historical SST data, it can predict the SST change over the next 50 years in the same region and quantify the impact of different RCPs indicators on SST changes. Through the simulation experiment, taking the North Atlantic region as an example, using the summer data of 1870-1966, the summer seawater temperature in 2016 was predicted. Through the cosine similarity test, the error cosine value between the expected result and the actual SST was 0.99984, and the Angle was 1.0115°. Under different RCPS standards, the average SST over the North Atlantic will reach 13.2℃, 13.8℃, 14.2℃ and 15.3℃, respectively, in summer 2070. The results show the irreversibility of SST rise and the severe influence of global warming and climate destruction on SST. The same simulation method is used to verify the prediction accuracy of the BP neural network, and the error Angle reached 2.2445°, which was much lower than the effect of RSST-ARGM. By comparison, RSST-ARGM has the advantages of long-term prediction ability, high prediction accuracy, low computational complexity, good robustness, and strong adaptability,

which can provide reliable data support for Marine research and environmental governance in relevant areas, and has a high practical value.

2 Related work

The ocean is one of the most critical areas to maintain the earth's ecology, in which abundant natural resources provide support for human life and production. The increasing severity of global warming leads to the irreversible rise of sea surface temperature, which has a significant impact on marine organisms and the physical and chemical indexes of regional seawater. The trend of sea warming will cause future rise, and it is difficult for the economic, social, and other fields to sustain growth. Therefore, the trend of ocean temperature change in the future will play an essential role in environmental protection and industrial development.

In recent years, many scholars around the world have used mathematical models to predict sea temperature trends. Mary-Louise (2018) studied the predicted results of the Arctic Ocean in the past 30 years and found that the ocean heat content nearly doubled. This warming was related to the abnormal heating of local surface water by the sun. The research results found that the heat absorbed by the local basin edge would accumulate in the ocean interior, causing the rise of sea temperature[1]. Zheng (2020) proposed an algorithm instead of the mathematical, physical model to predict the seawater temperature field by combining satellite data and a deep learning model[2]. Ratnam (2020) verified the superiority of the ANN algorithm in predicting Indian Ocean dipoles[3]. Hervieux (2019) proposed an anomaly prediction assessment method based on NMME for large Marine ecosystems off the coast of the United States and Canada, with a leveled approach to monthly SST to improve the overall prediction[4]. Qian (2020) compares the prediction effect of the statistical model of SST and the dynamic global circulation model on the seasonal precipitation in the Yangtze River Basin, finding that the statistical model had higher prediction performance, especially in long time span[5]. Dias (2019) adopts the inverse linear statistical model to forecast the sea temperature and sea temperature changes in the North Pacific Ocean, which is better than NMME in seasonal forecasting ability[6]. Sohn (2016) used the multi-mode integration method to predict the accuracy of seawater temperature. The ENSO intensity had a serious impact on the accuracy of seawater temperature prediction, indicating that the ENSO prediction model was not fully applicable to the global SST prediction[7]. Capotondi (2019) explored the influence of ENSO and local SST on USWC, and proposed a sensitivity model analysis method. They found that tropical SST anomalies have a significant impact on USWC, which can increase the predictability of anomalous SST[8]. Ionita (2020) studied low flow in the Rhine and Elbe river basins in Europe in summer and could use historical SST, sea level pressure, precipitation and other environmental information to predict low flow[9].Taye (2021) analyzed the differences between SST drivers of the rainy season (July-September) in different regions of Ethiopia, and proposed an improved regional seasonal forecasting method based on the local topography and climatic environment[10]. Counillon (2021) used the Norwegian climate prediction model in different configurations to study the impact of climate bias in the tropical Atlantic Ocean on seasonal prediction. Combined with the coupling of NORCPM and ensemble Kalman filter, Counillon corrected

the SST field exchange seen in the atmosphere and the ocean, reduced the climate model bias between precipitation and SST, and improved the accuracy of seasonal prediction[11]. Kale (2020) used monthly temperature, evaporation and precipitation data as input and combined with a variety of statistical methods to build an adaptive neuro-fuzzy reasoning system for regional SST, and verified the accuracy of the prediction from statistical standards[12]. Hotta (2019) proposed a method to improve the prediction accuracy of EPS in JMA by seawater temperature perturbation. The experiment proved that the perturbation of SST did not affect the ensemble mean forecast quality[13]. Jacox (2019) used the NMME Global Climate Prediction System to evaluate the maximum prediction period of seasonal SST in the California Ocean Current System (CCS) for up to 4 months and found that ENSO has varying degrees of impact on the sustainability of SST prediction[14]. Narapusetty (2018) proposed an interactive global method (CFSIE) using climate prediction system models to integrate the initial perturbation states of different CFS. It was found that the application of noise reduction in this method would reduce the ability of SST prediction[15]. Newman (2017) discussed the performance improvement space of the forecast model, compared the difference between the NMME results and the LIM results, and found that the prediction ability of the two models was close to the potential limitations. The study found large differences in some areas, but similar results in most areas[16]. Aurélien (2021) designed a new climate model to reduce the uncertainty of SST warming estimates by improving observational and statistical methods[17]. Francisco (2017) found that the areas most severely affected by global warming will experience an increase in regional water temperature dispersion, a gradual slowing down of water circulation, and a continuous decline in regional production capacity, by studying the reason why the location of the special biodiversity areas overlaps with the areas with severe global warming[18]. Ham (2019) used the transfer learning method to train the CNN neural network algorithm and improve its prediction time length for the ENSO phenomenon. Through the training and verification of historical data, the prediction accuracy is due to the traditional model prediction effect[19]. Sévellec (2018) found that the chaos of the climate system would affect the accuracy of the prediction model. Then he designed a probabilistic prediction method based on the transfer operator, which has a good prediction effect on the global average temperature and SST[20]. Chikamoto (2020) proved that a prediction system with fully coupled climate models could predict the annual water supply to the Colorado River years in advance. The model result showed that chronic water shortages in the Colorado River are significantly correlated with precursors of sea surface temperatures, including cooling in the tropical Pacific, warming in the northern Pacific and warming in the southern tropical Atlantic[21].Hermanson (2017) studied the drift detection of SST and precipitation by two seasonal forecast systems, and also asked that the drift was different between different forecast systems, which would produce phenomena such as overdraft and inverse drift, thus affecting the accuracy of model forecast[22]. Seager (2019) found that the concentration of greenhouse gases in nature for the sharp rise in a short time, caused by the west to the east of cold and warm to the strengthening of the surface temperature gradient, and the current model have obvious difference, through the introduction of atmospheric

dynamics and thermodynamics of regression analysis can reduce the deviation of parallel to the surface. However, in some areas, the warming trend is still hard to predict, especially the SST in the sensitive region of serious distortion prediction results[23]. Shaltout (2019) studied the history of the red sea surface temperature data, found in the spring and autumn season, seawater temperature gradient is higher than in the summer and winter season. Mean sea level pressure, elevation, and the temperature are major influential factors of sea surface temperature. At the same time, Shaltout explore the effect of different carbon dioxide concentrations in the future temperature changes, the GFDL - CM3 method of sea surface temperature change is forecasted, found the red sea will experience a significant increasing trend[24]. Perrie (2017) discussed how the climate changes influence the SST in the Barents Sea, showing that the estimation of local SST and ice volume by traditional model simulation was biased to a certain extent, so it could not accurately reflect the influence of heat on SST. By introducing a new model to calculate the contribution of solar radiation to SST change, and through historical data prove that heat transport and solar radiation are the main reasons for the increase of local SST, so as to predict the trend of SST change from 2010 to 2040[25]. Chris (2018) using multiple regression rebuilt the GST sequence since 1891, compared with CMIP5 model obtained the different force factor's contribution to the GST change, mostly was caused by the increase of greenhouse gases and artificial aerosol, weakened the solar radiation can produce cool, but for the near region of sea surface temperature forecast will produce greater uncertainty[26]. Melissa (2018) pointed out that the warming of North Atlantic SST is one of the significant features predicted by global climate models. Through the simulation analysis of the large set of community earth system models, he explained the relationship between the development of warm hole and AMOC, and studied the influence of the change of North Atlantic regional ocean current on the change of local SST, finally found an increasing ocean advective heat flux divergence within the center of the subpolar gyre, causing this warming deficit in the SST of North Atlantic, causing the slowdown of AMOC[27]. Ogurtsov (2016) analyzed regional summer temperature, North Atlantic SST and solar activity during 1567-1986, and the study showed that the solar activity and regional summer temperature were significantly correlated with North Atlantic SST during 1715-1896, and the change of SST may be the physical factor that transferred the influence of the sun on the regional temperature[28].

It can be seen from the existing studies that the main research direction is to use SST as a meteorological element to make short-term prediction of climate phenomena such as drought, ENSO, hardness like a dipole and regional water flow, and global warming will affect regional SST.

Obviously, these predictions just tried to build a relationship between the SST and other climates, and no research has shown that the two factors cause and affect each other[29]. Thus, before determining the influence of climatic factors on SST, it has nothing to do with studying the changes of SST.

There are relatively a few studies on the prediction of SST changes, most of which use neural networks, classical climate models or mathematical statistical methods to make short-term predictions of SST changes within 6-12 months. As for the most popular neural network methods, simply applying the existing model methods may

not achieve better experimental results. Chen(2020) applied deep learning to traffic cloud computing, especially improved the ability of extracting feature information from the original training model, and improved the detection effect and accuracy of equipment[30]. Zhao(2019) pointed out that machine learning alone cannot achieve ideal results on all occasions. He proposed an optimization analysis method that reduces fuzzy bias and improves the accuracy and flexibility of deep learning[31]. Gao(2020) systematically evaluated the performance of deep learning features in model retrieval, and compares hand-crafted features and deep learning features in 3D model retrieval. Thus, the robustness and computational complexity of this deep learning algorithm were verified[32]. And that's clear that the results only obtained by applying different neural network methods are still limited: Neural networks and climate models have good prediction accuracy in the short term through big data training and adding variables[33].

However, if the same method is used to apply deep learning to the long-term prediction, the amount of data in the training set will continue to climb with the increase of the prediction time range. Deep learning studies in other fields have found that the robustness and processing power of the model under large data volumes are not as good as the effect of short-term prediction[34]. The model must be modified according to the characteristics of the data set[35], but this also reduces the portability of the model.

What's more, mathematical statistical methods have vital requirements on sample distribution law and lack the ability to deal with abnormal situations, resulting in disconnection with the actual situation, What's more, mathematical statistical methods have substantial requirements on sample distribution law and they lack the ability to deal with abnormal situations, resulting in disconnection with the actual conditions[36]. Simultaneously, the historical SST data set used in short-term prediction is only a few months to a year, so the data volume is too small to provide long-term data support for SST changes, long-time prediction with small amount of data will reduce the accuracy of prediction results. Besides, few studies have explored the relationship between SST changes from the perspective of global warming, ignoring the effect of climate, one of the most important factors, on SST changes.

3 Methods/Experimental

The aim of this paper is to solve the problem for regional SST prediction in long time span by a newly designed gray prediction model based on atmospheric reflection.

Firstly, the steps of introducing the grey model and getting the prediction result sequence by using the regional historical SST are introduced. Then a quantitative model of the impact of different greenhouse gas emission scenarios on future SST changes is established. Finally, the North Atlantic Ocean was taken as the simulation object to predict the local SST in 2016 according to the local data from 1870 to 1966, to verify the effectiveness of the method, and to predict the SST in this region in the next 50 years under different greenhouse gas emission scenarios.

4 Preliminary Work

These problems above are the characters of SST prediction, this paper attempts to build a model with the ability to predict regional long-term sea temperature

change[37], so that it can have a high prediction accuracy in a long-time span and investigate the impact of global warming on SST change, which provides data support for improving people's understanding of sea water temperature change law and promoting the sound development of economy and society. Because the ocean is influenced by many factors, including climate, region, and human activities, the causes of SST changes are very complex and it is difficult to quantify the influence of different factors on SST. Due to the complexity of SST change, its changing progress from the outside is a gray system. This paper hopes to propose a method to reveal the characteristics and degree of SST change, which is a typical gray model application scenario, and it can be used to predict SST based on the gray model method.

4.1 Grey Model

Grey model is established based on grey system theory. When the hierarchical or structural relationship of a system is fuzzy, the dynamic change is random, and the index data is incomplete or uncertain, these characteristics are all called grayness, and the system with grayness is a gray system. Grey system contains specific information and uncertain information, so the grey model is insensitive to data regularity, information content and information integrity. The grey model is able to forecast the changing tendency of the object of observation from the system information[38]. After processing the discrete data series, the randomness of the model is significantly reduced and it becomes a regular generation number. Therefore, the dynamic model in the form of differential equation is constructed, which has approximation and non-uniqueness.

For the forecasting of SST, we use the gray prediction model, which helps us analyze the correlation between historical seawater temperature observation data and infer the future seawater temperature trend through the dynamic change of historical water temperature. After the existing temperature data is quantified, the generated number with strong regularity is obtained by the cumulative generation method, and then the differential equation model GM (1,1) is obtained through the function change. The gray prediction is to predict the development law of the seawater temperature system using our seawater temperature grey model.

4.2 RCPs

Global climate model is the primary tool for predicting climate changes, studying climate change and its response. The Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR5)[39], based on the CMIP5 model, has carried out studies from model evaluation to future climate prediction of several global climate models, and provided source data for high-resolution regional climate change studies[40]. CMIP5 introduces a new greenhouse gas emission scenario based on CMIP3, named RCPs. Representative Concentration Pathways (RCPs) are a set of greenhouse gas emission scenarios set in global climate models for the study and prediction of future climate change[41]. It mainly reflects the situation of carbon dioxide emissions and the impact of different levels of carbon dioxide emissions on the Atmospheric Radiation.

By studying the relationship between RCPs and SST changes, the influence of global warming on SST can be established. RCPs mainly target four levels of greenhouse gases: 2.6, 4.5, 6.0, and 8.5. RCP8.5 stands for the highest level of carbon dioxide emissions, or as much carbon dioxide as possible; RCP2.6 represents the minimum level of carbon dioxide emissions, which means that human society can emit as little or no carbon dioxide as possible under the current production conditions. RCP4.5 and RCP6.0, which are between RCP8.5 and RCP2.6, are considered as the most likely carbon emission scenarios in the future.

5 RSST-ARGM: a Regional Sea Surface Temperature - Atmospheric Radiation Grey Model

The following conclusions can be drawn from the study of SST changes: SST changes are related to past sea surface temperatures, sea surface temperature changes are affected by many factors, sea surface temperature trends exist regionally and global warming is one of the main reasons that affect sea water temperature rise. So that we can conclude that the changing progress of SST conform to the grey system, and the prediction model has regional restriction, which means the prediction model has better performance in the specific area than that in the whole world. And the different degrees of global warming in the future will also affect the degree and speed of SST change.

Based on the above method, a grey prediction model of regional sea surface temperature based on atmospheric reflection is established according to the characteristics of SST, named regional sea surface temperature-atmospheric radiation grey model (RSST-ARGM).

In the process of building the model, several steps are needed, including data preprocessing, grey system of SST and influence of future atmospheric reflection.

5.1 Data checking and preprocessing

According to the requirements of the model, the seawater surface temperature data needs to be preprocessed, in which the defects of the data set itself should be properly cleaned to ensure its integrity and effectiveness. In addition, the gray system itself has some requirements for the data sets.

SST data are collected from observed historical record, and the main possible problems are abnormal data, namely data missing and incorrect statistics.

Data cleaning is needed for abnormal data. Delete or fill in the missing data, which depends on the proportion of the accurate data. In this paper, it is stipulated that if the missing rate of the data set is less than 5 percent, the data will be deleted, that is, the data of the month or year will not be counted, to prevent the artificial influence on the data. If the missing rate is greater than or equal to 5 percent, the lost data will be filled by means of taking the mean value of the two data before and after the missing data, to avoid the prediction error caused by too much indeed.

The primary index data in the wrong statistics show obviously unreasonable values, because the sea water temperature changes relatively gently, and the temperature in a region generally fluctuates within a relative range.

Therefore, the error statistics were checked by using box plot: the data of every day in a fixed position in a year were plotted using box plot. If several points were

too high or too low, statistical errors at that point would be considered and treated as missing values. After processing the abnormal data, the feasibility of the grey system model is needed to be ensured, the data need to be checked and processed. The main steps are as follows:

1. Set the reference sequence be

$$a^{(0)} = (a^{(0)}(1), a^{(0)}(2), ..., a^{(0)}(n)) \tag{1}$$

2. Calculate the series ratio

$$\lambda(s) = \frac{a^{(0)}(s-1)}{a^{(0)}(s)}, s = 2, 3, ..., n \tag{2}$$

3. If every step ratio $\lambda(k)$ is within the tolerable coverage $\theta = (e^{(-\frac{2}{n+1})}, e^{(\frac{2}{n+2})})$, then $x^{(0)}$ can be performed as data of Gray model GM (1,1). Otherwise, we need to perform a translation transformation on $x^{(0)}$,

$$y^{(0)}(s) = a^{(0)}(s) + c, s = 1, 2, ..., n \tag{3}$$

The sequence $y^{(0)} = (y^{(0)}(1), y^{(0)}(2), ..., y^{(0)}(n))$ is

$$\lambda_y(s) = \frac{y^{(0)}(s-1)}{y^{(0)}(s)} \in \theta, s = 2, 3, ..., n \tag{4}$$

5.2 Grey system prediction of SST

In the grey model, the original sequence (0) offers the original data and generate a new row (1) through accumulation and other changes, to make the randomness of the original data weakly and exposure it hidden characteristic rules. A differential equation model was established for the generated sequence (1)[42].

The GM(1,1) model used in this paper represents the differential equation model of 1 order and 1 variable.

The data set after data inspection and preprocessing meets the requirements of the gray system, and the data is put into the gray system through the following steps:

1. Accumulate a reference data column

$$a^{(0)} = (a^{(0)}(1), a^{(0)}(2), ..., a^{(0)}(n))$$

At one time to calculate a new series

$$a^{(1)} = [a^{(1)}(1), a^{(1)}(2), ..., a^{(1)}(n)] = [a^{(1)}(1), a^{(1)}(1)+a^{(0)}(2), ..., a^{(1)}(n-1)+a^{(0)}(n)]$$

among the series, $a^{(1)}(s) = \sum_{i=1}^{s} a^{(0)}(i), s = 1, 2, ..., n$

2. Calculate the mean number of the series

$$z^{(1)}(s) = 0.5a^{(1)}(s) + 0.5a^{(1)}(s-1), s = 2, 3, ..., n \tag{5}$$

So

$$z(1) = (z^{(1)}(2), z^{(1)}(3), ..., z^{(1)}(n))$$

The differential equation is

$$a^{(0)}(s) + xz^{(1)}(s) = b, s = 2, 3, ..., n \tag{6}$$

The corresponding albino differential equation is

$$\frac{da^{(1)}}{dt} + xa^{(1)}(t) = b.$$

3. Mark related parameter

$$u = (x, b)^T, Y = \left(a^{(0)}(2), a^{(0)}(3), ..., a^{(0)}(n)\right)^T, b = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$$

Obtained by least squares $\hat{u} = (x, b)^T = (B^T B)^{(-1)} B^T Y$, which make $J(\hat{u}) = (Y - B\hat{u})^T (Y - B\hat{u})$.

The result of the corresponding albino differential equation is

$$\hat{a}^{(1)}(s+1) = (a^{(0)}(1) - \frac{b}{x})e^{-xk} + \frac{b}{x}, s = 0, 1, ..., n-1, ... \tag{7}$$

among them, $\hat{a}^{(0)}(s+1) = \hat{a}^{(1)}(s+1) - \hat{a}^{(1)}(s), s = 1, 2, ..., n-1, ...$

4. Test the predicted value According to the model requirements, if $\varepsilon(s) < 0.2$, $\varepsilon(s) = [a^{(0)}(s) - \hat{a}^{(0)}(s)]/a^{(0)}(s)$,$s = 1, 2, ..., n$, $\hat{a}^{(0)}(1) = a^{(0)}(1)$, the predicted results meet the requirements. And then, calculate $\lambda(k)$ according to $a^{(0)}(s-1)$ and $a^{(0)}(s)$. Use coefficient an in **equation**(6) to Calculate $\rho(s) = 1 - [(1 - 0.5x)/(1 + 0.5x)] \times \lambda(s)$. If $\rho(s) < 0.2$, it meets the requirement. If the $\varepsilon(s)$ nor $\rho(s)$ meet the requirements, we are required to readjust the data set and the corresponding parameters until the results meet inspection standards.

5.3 RCPs

The temperature changes under different RCPS indexes are very other. Try to quantify the effect of RCPS index on temperature change, we need to divide the existing data set into prediction data set and detection data set. The influence factors of RCP2.6, 4.5, 6.0 and 8.5 are defined as:

$$\gamma_i = \frac{t_r}{t_i}, i = 1, 2, 3, 4$$

among them, t_i is the temperature predicted by using the prediction data set under the $i - th$ RCP condition, t_r is the actual current temperature.

On that basis, the final predicted SST sequence is:

$$T(k+1) = \gamma_i \times \hat{x}^{(1)}(k+1), k = 1, 2, ..., n-1, ...$$

6 Result

Theoretical derivation alone is not enough, the RSST-ARGM prediction model is required be proved to have high prediction accuracy, a simulation experiment is needed. There is the world-famous North Sea fishing ground in the North Atlantic, and many neighboring countries have developed advanced marine fishery with good natural environment. The future change of this sea area is of great significance in ecological, environmental protection and social and economic aspects[43]. Thus, take the sea area of the North Atlantic (from 14.5°w to 10.5°e, and 45.5°n to 65.5°n) as an example, as is shown in **Fig.1**, explore local sea surface temperatures over the next 50 years.

6.1 Data source

The data set used for model simulation in this paper is Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST), which is derived from the Met Office Hadley Centre observations datasets. This data set records surface temperatures in all waters from 1870 to November 2020[44]. This data set maintains high reliability in temperature statistics and equipment updates over the years, and can be used in professional data research[45]. The location data of the North Atlantic Ocean are selected in the data set and brought into the model for simulation.

6.2 Simulation design

The simulation predicted the North Atlantic SST, and the geographical location represented by the data was between 14.5°W to 10.5°E, and 45.5°N to 65.5°N, and the data set was from January 1, 1870 to December 31, 2020.

In order to improve the test efficiency, we mainly investigate the 2 months when the sea temperature is the highest. The forecast is for the next 50 years, from 2021 to 2070.

6.3 Simulation result

Because of the large amount of data, it is obviously difficult to test each data result. Therefore, in this paper, the data from 1870 to 1966 are used as the model input to predict 50 years later, namely, SST in 2016 and compare it with the actual SST measured in 2016. In order to restore the real situation as much as possible, RCP is set to 6.0. The predicted result of the model is the temperature of each point in this area, which is an extensive three-dimensional data set. In order to compare the products more intuitively, the temperature point set is presented in the form of heat map. Set the same temperature ruler as the actual temperature chart, and get two heat maps of the same region (from 0 to 14.5° W, and 45.5° N to 63.5° N) successively, as is shown in **Fig. 1**. Using the Cosine Similarity Algorithm to compare the differences between two pictures[46]. According to the theory of Cosine Similarity Algorithm. If the cosine value is close to 1, the angle is relative to 0 ° (the coincidence state), the two vectors have high similarity. Conversely, the more immediate to -1, the more significant the difference.

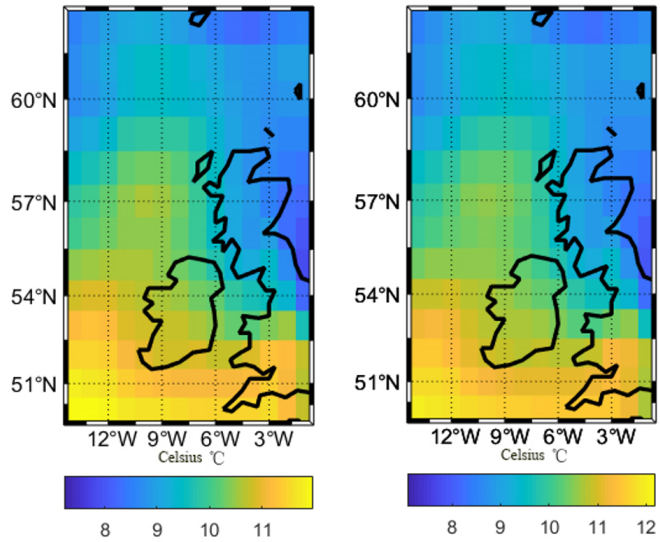

Figure 1 The results of model (left is the real situation; right is the model result)

The cosine similarity algorithm uses the cosine value between two vectors, and uses the Euclidean dot product formula to obtain the cosine similarity θ of two given attribute vectors A and B from the dot product and the vector length:

$$S = \cos(\theta) = \frac{A \cdot B}{||A|\,|B||} = \frac{\sum_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \cdot \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{8}$$

where A_i, B_i represents the components of the vectors A, B.

The test results show that the vector cosine of the sea temperature map based on the model's predicted sea temperature and the actual temperature is 0.99984, and the cosine angle is 1.0115° at the resolution of 1 longitude times 1 latitude. Prove the validity of data prediction.

Based on the RSST-ARGM prediction model, we obtained seawater temperature pictures 50 years later (2070) in the situation of RCP6.0, as is shown in **Fig. 2**.

It can be seen from the model results that the June-August sea temperature in the North Atlantic region (from 14.5°W to 10.5°E, and 45.5°N to 65.5°N) will generally rise in 2070, and the highest sea temperature in the region will reach 14.2℃, which is 1.7℃ higher than the maximum temperature 50 years ago. What's more, warming is more pronounced at high latitudes, resulting in the spread of hot areas.

In addition, as can be seen from the forecast map, the boundary between SST and land is becoming increasingly clear, especially in the western hemisphere. This may be because the waters in the Western Hemisphere are more integrated and less affected by ground than in the Eastern Hemisphere. This is further evidence that the range of ocean warming is increasing, leading to closer heat exchange between the oceans, which may further contribute to future warming.

Seawater temperature rises to a certain extent in the short term will increase local warm, Marine resources. Still, in the long term this will no doubt cause more

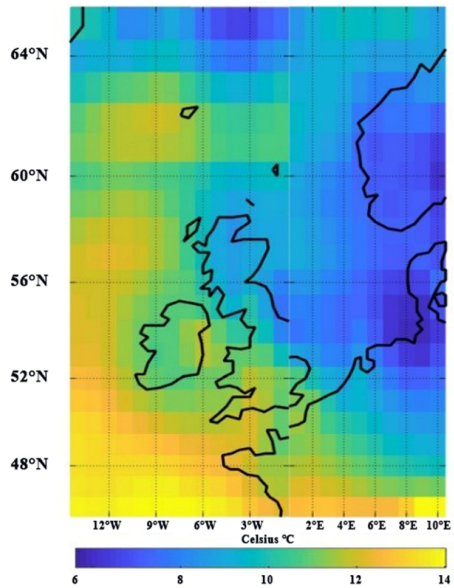

Figure 2 The regional SST in 2070

significant damage to the ecological balance of the whole, from existing research areas of water temperature rise will be found in seawater nutrient loss gradually, eventually make the production capacity of the water drops in the region until lost.

At the same time, the rise of sea temperature and the spread of high temperature range will further lead to the melting of glaciers, so as to raise the sea level and cause natural disasters.

In this case, the physical and chemical indexes of seawater will change to some extent, and the impact is difficult to estimate.

From the perspective of average SST, no matter what RCPs index is, SST keeps growing during 2021-2030, and the temperature of growth is similar to that of about 13℃. After 2030, the seawater temperature at RCP2.6 tends to be stable, and will remain at 13.2℃ in 2070. RCP4.5, RCP6.0 and RCP8.5 increased continuously, reaching 13.8℃, 14.2℃ and 15.3℃ respectively in 2070, as is shown in **Tbl. 1**. The comparation of temperature changing trend is shown in **Fig. 3**.

The data of the four prediction results have vividly shown that the higher the RCP index is, the more pronounced the growth of SST will be, indicating the influence degree of atmospheric radiation on SST change under the influence of global warming.

Table 1 Prediection result in 2017 under different RCPs

RCP	temperature(℃)	trend
2.6	13.2	stable
4.5	13.8	rise slowly
6.0	14.2	rise
8.5	15.3	rise sharply

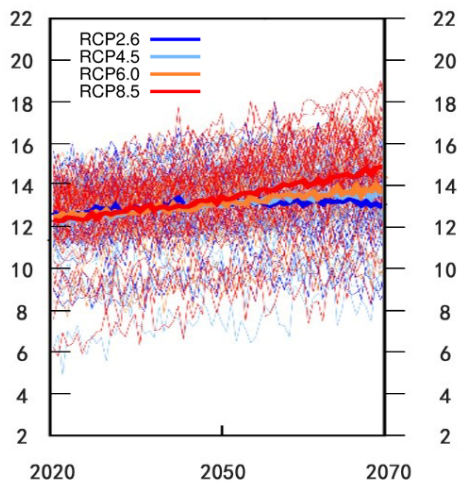

Figure 3 The results of SST changes under the influence of different RCPs

It is worth noting that even the predicted SST under RCP2.6 does not show an apparent downward trend after rising to 13.2℃, which means that the warming of SST is almost an irreversible process under the advanced technical means.

Strict control of greenhouse gas emission and environmental pollution can slow down the trend and speed of warming. If the arbitrary emission and destruction will make the sea surface temperature rise to 15.3℃, the damage to the global climate is unimaginable.

7 Discussion

In order to further evaluate the performance of the RSST-ARGM prediction model, this paper attempts to compare the prediction results of the RSST-ARGM model in the North Atlantic with the neural network method commonly used in the current SST prediction, and to show the characteristics of the proposed method.

This comparison is using the most classic BP neural network, still using the data from 1870 to 1966 are used as the model input to predict 50 years later, namely, SST in 2016 and compare it with the actual SST measured in 2016[47]. Considering that the neural network model needs to be trained in advance, the eastern hemisphere part data of the North Atlantic(from 0 to 10.5°E, and 45.5°N to 65.5°N) are selected as a training set, and the data of the western part(14.5°W to 0, and 45.5°N to 65.5°N) are used as the prediction result, and are compared with the products of RSST-ARGM.

In the BP network, data from 1870-1966 are used as the input layer, and temperature data from 2016 are used as the output layer. Input layer points are 96, output layer points are 1. According to the empirical formula of hidden layer neuron , the number of neurons are among 11 to 20[48]. In this paper, we choose 14 as the neuron number. The excitation functions of hidden layer and output layer, which have great importance in the performance of the system, are set as tansig and logsig functions respectively, traindx is setted as the network training function, and the

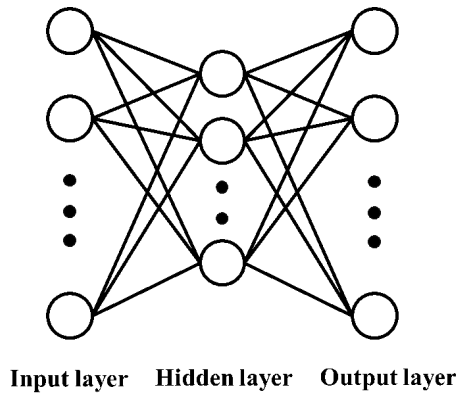

Figure 4 BP neural network structure diagram

network iterations number matters a lot setting as 5000. To get good results, the expected error is , the rate of model learning is 0.01[49], as is shown in **Fig.4**.

The test set data is brought into the model, and the cosine similarity method is also used to get the predicted results and actual data. The average error is 0.99923, convert into angel is 2.2445°, its accuracy is significantly lower than that of RSST-ARGM (the neural network results are different each time, repeat 5 times to take the average value). The comparison of the prediction effects of the two models is shown in **Tbl. 2**

Table 2 Comparison of the prediction effects of the two models

	training set	prediection set	angle(°)	vector
RSST-ARGM	1870-1966	2016	1.0115	0.99984
BP network			2.2445	0.99923

If the accuracy of the neural network is to be further enhanced, the training times and expected error accuracy can only be improved based on unchanged data set, which will significantly prolong the training time of the model and increase the test cost.

Compared with neural network prediction, the result can be clearly proved that the prediction accuracy of RSST-ARGM model is high, and the requirement of computing ability is low. In addition, the introduction of RCPS index can further reflect the extent to which global warming changes atmospheric radiation and thus affects SST, which can provide more scientific data support for the long-term study on dynamic changes of SST.

8 Conclusion

In this paper, we combine the grey model system and Representative Concentration Pathways in Climatology, establishing a gray prediction model based on regional historical SST and atmospheric reflection conditions, RSST-ARGM. Based on the calculation of historical SST, the model explores the trend of SST change in the future, and introduces the influence of atmospheric greenhouse gas concentration change on SST change in the form of RCPs index, to reflect the correlation between global warming and SST change. Through simulation, in this article, the

RSST–ARGM model is applied to predict the temperature changes of the next 50 years of the North Atlantic waters. By contrast with classical neural network model, the result manifests the RSST-ARGM model is easy to calculate, with advantages of the high accuracy and good robustness. The model's prediction data results reveal a rise in sea surface temperatures in the future of sustainability and diffusivity and the promoting effect of global warming on sea surface temperature rising (different RCPs index has had a huge impact on overall temperature).

The prediction model constructed in this paper has high accuracy over a 50 year time span. However, in order to simplify the calculation, historical SST data were only used as much as possible in this model, while the influence of different historical data capacity on the accuracy of the results was not studied. Since the trend of SST change and global warming in different stages in the past are not the same, the RCPs indicators corresponding to various historical data stages can be further explored in the future, to further improve the pertinence and accuracy of the prediction. Besides, the relationship between climate and seawater temperature can be further quantitatively analyzed supported by the results of the model and the knowledge of seawater physical and chemical composition and thermodynamics.

Abbreviations

Subject noun

These abbreviations are the nouns of the subject area:

SST – Sea Surface Temperature
RCPs – Representative Concentration Pathways
NMME – North American Multi-Model Ensemble
ENSO – El Niño and La Niña
USWC – the US West Coast
NorCPM – the Norwegian Climate Prediction Model
EPS – Ensemble Prediction System
JMA – the Japan Meteorological Agency
CCS – California Ocean Current System
CFS – Climate Forecasting System
CFSIE – CFS in an Interactive Ensemble approach
GFDL – Geophysical Fluid Dynamics Laboratory
GST – Global mean Surface Temperature
AMOC – the Atlantic Meridional Overturning Circulation
IPCC AR5 – the Fifth Assessment Report of the Intergovern-mental Panel on Climate Change
CMIP3/5 – Coupled Model Intercomparison Project

Algorithm

These abbreviations are the nouns of mathematical algorithm:

GM – Grey Model
ANN – Artificial Neural Network
CNN – Convolutional Neural Network
RSST-ARGM – a Regional Sea Surface Temperature-Atmospheric Radiation Grey Model

Source of data set

This abbreviation is the providing unit of the data set

HadISST – Hadley Centre Sea Ice and Sea Surface Temperature

Declarations

Availability of data and materials

The data set used for model simulation in this paper was supplied by Hadley Centre Sea Ice and Sea Surface Temperature data set (HadISST)

Competing interset

The authors declared that they have no conflicts of interest to this work.

Funding

This work was supported in part by the National College Students Innovation and Entrepreneurship Program (No. 202010300061), Jiangsu College Students Innovation and Entrepreneurship Program (No. 202010300008Z), "Research on the informationalized construction of emergency management system from the perspective of big data", National Natural Science Foundation of China (No. 41911530242, 41975142, 41875184), 5150 Spring Specialists (05492018012, 05762018039), Major Program of the National Social Science Fund of China (Grant No.17ZDA092), 333 High-Level Talent Cultivation Project of Jiangsu Province (BRA2018332), Royal Society of Edinburgh, UK and China Natural Science Foundation Council (RSE Reference: 62967_Liu_2018_2) under their Joint International Projects funding scheme, Innovation Team of "Six Talent Peaks" in Jiangsu Province (Grant No.TD-XYDXX-004) and basic Research Programs (Natural Science Foundation) of Jiangsu Province (BK20191398, BK20180794).

Authors contribution

LZ finished the algorithm and the writing of this paper. QL designed the progress of simulation. XL put forward the idea of this paper. YZ analyzed relevant studies and proposed improved methods. All authors read and approved the final manuscript.

Acknowledgements

In the writing of this paper, we would like to thank Professor Wenjun Liu of Nanjing University of Information Science and Technology. For his advice and theoretical guidance on the research topic and research direction. With his help, our paper has a good academic and scientific performance. Yanli Chen, associate professor of Nanjing University of Information Science and Technology, is the instructor of the National College Students Innovation and Entrepreneurship Program project. She provided us with ideas for the writing of this article and provided a lot of help for the smooth progress of the writing of this article.

Authors' information

Zhu Linqian, is a B. Sc. candidate at Nanjing University of Information Science and Technology. His main research interest include data structure, and digital signal detection and processing.

Author details

[1]School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China. [2]Scholl of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China. [3]Centre for Information and School of Computing, Edinburgh Napier University, Scotland, UK. [4]School of Automation, Nanjing University of Information Science and Technology, Nanjing, China.

References

1. Mary-Louise, T., John, T., Richard, K.: Warming of the interior arctic ocean linked to sea ice losses at the basin margins. ence Advances **4**(8), 6773 (2018)
2. Zheng, G., Li, X., Zhang, R.H., Liu, B.: Purely satellite data–driven deep learning forecast of complicated tropical instability waves. Science Advances **6**(29), 1482 (2020)
3. Ratnam, J.V., Dijkstra, H.A., Behera, S.K.: A machine learning based prediction system for the indian ocean dipole. Scientific Reports **10**(1), 284 (2020)
4. Hervieux, G., Alexander, M.A., Stock, C.A., Jacox, M., Tommasi, D.: More reliable coastal sst forecasts from the north american multimodel ensemble. Climate Dynamics, 1–16 (2012)
5. S.Qian, J.Chen, X. Li, e.a.: Seasonal rainfall forecasting for the yangtze river basin using statistical and dynamical models. International Journal of Climatology **40**(1) (2020)
6. Dias, D.F., Subramanian, A., Zanna, L., Miller, A.J.: Remote and local influences in forecasting pacific sst: a linear inverse model and a multimodel ensemble study. Climate Dynamics (2019)
7. Sohn, S.J., Tam, C.Y., Jeong, H.I.: How do the strength and type of enso affect sst predictability in coupled models. Scientific Reports **6**, 33790 (2016)
8. Capotondi, A., Sardeshmukh, P.D., Lorenzo, E.D., Subramanian, A.C., Miller, A.J.: Predictability of us west coast ocean temperatures is not solely due to enso. Scientific Reports **9**(1) (2019)
9. Ionita, M., Nagavciuc, V.: Forecasting low flow conditions months in advance through teleconnection patterns, with a special focus on summer 2018. Scientific Reports

10. Mtt, A., Ed, B., Kjc, B., Lch, C.: Potential predictability of the ethiopian summer rains: Understanding local variations and their implications for water management decisions. Science of The Total Environment (2020)
11. Counillon, F., Keenlyside, N., Toniazzo, T., Koseki, S., Wang, Y.: Relating model bias and prediction skill in the equatorial atlantic. Climate Dynamics, 1–14 (2021)
12. Kale, S.: Development of an adaptive neuro-fuzzy inference system (anfis) model to predict sea surface temperature (sst). Oceanological and Hydrobiological Studies **49**(4), 354–373 (2020)
13. Hotta, D., Ota, Y.: Statistical generation of sst perturbations with spatio-temporally coherent growing patterns. Quarterly Journal of the Royal Meteorological Society **145** (2019)
14. Jacox, M.G., Alexander, M.A., Stock, C.A., Hervieux, G.: On the skill of seasonal sea surface temperature forecasts in the california current system and its connection to enso variability. Climate Dynamics (2017)
15. Narapusetty, B.: The role of atmospheric internal variability on the prediction skill of interannual north pacific sea-surface temperatures. Theoretical and Applied Climatology **133**(11), 1–9 (2017)
16. Newman, M., Sardeshmukh, P.D.: Are we near the predictability limit of tropical indo-pacific sea surface temperatures? Geophysical Research Letters (2017)
17. Ribes A., Q.S., P., G.N.: Making climate projections conditional on historical observations. Science Advances **7**(4), 0671 (2021)
18. Francisco, R., Isabel e.t., a.: Climate impacts on global hot spots of marine biodiversity. Science advances (2017)
19. Ham YG., K.J., JJ, L.: Deep learning for multi-year enso forecasts. Nature **573**, 568–572 (2019)
20. Sévellec, F., Drijfhout, S.S.: A novel probabilistic forecast system predicting anomalously warm 2018-2022 reinforcing the long-term global warming trend. Nature Communications **9** (2018)
21. Chikamoto, Y., Wang, S.-Y., Yost, M., Yocom, L., Gillies, R.: Colorado river water supply is predictable on multi-year timescales owing to long-term ocean memory. Nature Reviews Earth and Environment **1** (2020). doi:10.1038/s43247-020-00027-0
22. Hermanson, L., Ren, H.L., Vellinga, M., Dunstone, N.D., Hyder, P., Ineson, S., Scaife, A.A., Smith, D.M., Thompson, V., Tian, B.: Different types of drifts in two seasonal forecast systems and their dependence on enso. Climate Dynamics (2017)
23. Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathey, R., Zhang, H.: Strengthening tropical pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. Nature Climate Change **9**, 517–522 (2019). doi:10.1038/s41558-019-0505-x
24. Shaltout, M.: Recent sea surface temperature trends and future scenarios for the red sea. OCEANOLOGIA (2019)
25. Long, Z., Perrie, W.: Changes in ocean temperature in the barents sea in the 21 st century. Journal of Climate **30**(15) (2017)
26. Folland, C.K., Olivier, B., Andrew, C., Parker, D.E.: Causes of irregularities in trends of global mean surface temperature since the late 19th century. Science Advances **4**(6), 5297 (2018)
27. Melissa, G., Jeffrey, S., Yochanan, K.: Mechanisms governing the development of the north atlantic warming hole in the cesm-le future climate simulations. Journal of Climate, 17–06351 (2018)
28. Ogurtsov, M., Lindholm, M., Jalkanen, R., Veretenenko, S.V.: North atlantic sea surface temperature, solar activity and the climate of northern fennoscandia. Advances in Space Research **59**(4), 980–986 (2016)
29. Wang, L., Zhen, H., Fang, X., Wan, S., Guo, Y.: A unified two-parallel-branch deep neural network for joint gland contour and segmentation learning. Future Generation Computer Systems **100** (2019)
30. Chen, C., Liu, B., Wan, S., Qiao, P., Pei, Q.: An edge traffic flow detection scheme based on deep learning in an intelligent transportation system. IEEE Transactions on Intelligent Transportation Systems **PP**(99), 1–13 (2020)
31. Zhao, Y., Li, H., Wan, S., Sekuboyina, A., Hu, X., Tetteh, G., Piraud, M., Menze, B.: Knowledge-aided convolutional neural network for small organ segmentation. IEEE Journal of Biomedical and Health Informatics, 1–1 (2019)
32. Gao, Z., Li, Y., Wan, S.: Exploring deep learning for view-based 3d model retrieval. ACM Transactions on Multimedia Computing Communications and Applications **16**(1), 1–21 (2020)
33. Kiyohara, S., Tsubaki, M., Mizoguchi, T.: Learning excited states from ground states by using an artificial neural network. npj Computational Materials
34. Wan, S., Xu, X., Wang, T., Gu, Z.: An intelligent video analysis method for abnormal event detection in intelligent transportation systems. IEEE Transactions on Intelligent Transportation Systems, 1–9 (2020). doi:10.1109/TITS.2020.3017505
35. Ding, S., Qu, S., Xi, Y., Wan, S.: Stimulus-driven and concept-driven analysis for image caption generation. Neurocomputing (2019)
36. Cuvelier, N., Bartell, S.M.: Shrinkage estimation of long-term water ingestion rates. Journal of Exposure Science and Environmental Epidemiology, 1–9 (2021)
37. Athanasiadis, P.J., Yeager, S., Kwon, Y.-O., Bellucci, D.W. Alessio Smith, Tibaldi, S.: Decadal predictability of north atlantic blocking and the nao. npj Climate and Atmospheric Science **3** (2020)
38. Edeling, W., Arabnejad, H., Sinclair, R., Suleimenova, D., Coveney, P.: The impact of uncertainty on predictions of the covidsim epidemiological code. Nature Computational Science **1**(2) (2021)
39. van Vuuren, D.P., Edmonds, J., Kainuma, e.a. Mikiko: The representative concentration pathways: an overview. Climatic Change **109** (2011)
40. Myhre G.and Boucher O.and FM Bréon, e.a.: Declining uncertainty in transient climate response as co2 forcing dominates future climatechange. Nature Geoscience **8**(3), 181–185 (2018)
41. J., R., Rao S., e.a.: Air-pollution emission ranges consistent with the representative concentration pathways. Nature Climate Change
42. X., H., J., C.: A hybrid prediction model based on improved multivariable grey model for long-term electricity consumption. Electrical Engineering (1), 1–13 (2020)
43. Wang, H., Zhang, J., Jin, W., Yan, Z.: Research on mangrove recognition based on hyperspectral unmixing. In:

2017 IEEE International Conference on Unmanned Systems (ICUS) (2017)
44. Lee, J., Wasserman, W.W., Hoffmann, G.F., Karnebeek, C., Blau, N.: Knowledge base and mini-expert platform for the diagnosis of inborn errors of metabolism. Genetics in medicine: official journal of the American College of Medical Genetics **20**(1), 151–158 (2018)
45. Hausfather, Z., Cowtan, K., Clarke, D., Jacobs, P., Richardson, M., Rohde, R.: Assessing recent warming using instrumentally homogeneous sea surface temperature records. Science Advances **3**(1), 1601207 (2017)
46. Rayner, N., A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. Journal of Geophysical Research: Atmospheres **108**(D14) (2003)
47. Niu, J., Shen, L., Zheng, Q.: A measurement method of software flexibility based on bp network. In: International Workshop on Intelligent Systems and Applications (2009)
48. Liu, Y., Fu, H., Zhou, Y.: Application of neural network model in insulating oil fault diagnosis. In: International Conference on Measuring Technology Mechatronics Automation (2010)
49. Yang, G., Yuan, X.C.: Bank customer classification model based on elman neural network optimized by pso. IEEE, 5667–5670 (2007)

Figures

Tables