

RESEARCH

Open Access



RSST-ARGM: a data-driven approach to long-term sea surface temperature prediction

Linqian Zhu^{1*}, Qi Liu², Xiaodong Liu³ and Yonghong Zhang⁴

*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 sea surface temperature (SST), this paper studies the historical SST in regional sea areas and the emission pattern of greenhouse gases, proposing a Grey model of regional SST 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 variation, and further introduces different greenhouse gas emission scenarios in the future as the indexes 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. The accuracy of the model is as high as 0.99984. The model predicts that the regional SST could reach a maximum of 15.3 °C by 2070. This model is easy to calculate, with advantages of the high accuracy and good robustness.

Keywords: Long-term prediction, Regional SST, Temperature variation, Grey model, Atmospheric reflection

1 Introduction

The ocean is one of the most critical areas to maintain the ecology of Earth, 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. Many experts and scholars have done a lot of research on the problem of SST prediction, and put forward many innovative methods, but basically they have encountered such problems as short prediction time range, low accuracy, high algorithm complexity, and poor adaptability. Moreover, these methods often do not directly reflect the impact of environmental factors such as global warming on temperature variation, resulting in a decrease in the reference value of the prediction results

For the reason that the characteristics of regional sea surface temperature (SST) variation accord with the Grey system, combining theories of the Grey model and RCPs indexes. The prediction model of regional SST under atmospheric radiation, a Regional Sea Surface Temperature-Atmospheric Radiation Grey Model (RSST-ARGM) is built. Based on historical SST data, it can accurately predict the SST variation over the next 50 years in the same region and quantify the impact of different Representative

Concentration Pathways (RCPs) indicators on SST variation, building the relationship between the environmental factor and the SST.

Through the simulation experiment, taking the North Atlantic region as an example, using the data of 1870–1966, the summer sea surface 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 error angle was 1.0115° . Under different RCPs indexes, the average SST over the North Atlantic will reach 13.2°C , 13.8°C , 14.2°C and 15.3°C , 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 data were 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 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.

The following sections are organized as follows:

Part 2 describes the current research on the prediction of sea temperature change and the use of the main prediction models, and analyzes the challenges faced by these methods.

Part 3 puts forward the requirements for the new prediction model according to these challenges. By analyzing the causes of SST variation, a new prediction model is designed, and the main principles used in the model are introduced.

Part 4 describes in detail the construction process of RSST-ARGM method and the related calculation derivation, constructing a complete prediction model.

In Part 5, a set of simulation experiments are designed to verify the accuracy of the model results and calculate the prediction results of regional SST for the next 50 years.

Part 6 compares the classical neural network prediction methods, and shows that RSST-ARGM has obvious advantages in time span, prediction accuracy, complexity and adaptability.

In the last part (Part 7), the effects, advantages and simulation experiments of RSST-ARGM method are summarized, and some thoughts on future optimization and improvement are put forward.

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 indices 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 variation 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. Timmermans (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 (2017) proposed an anomaly prediction assessment method based on NMME for large Marine ecosystems off the coast of the USA 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 (2020) 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 (2017) 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 (2017) 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 explores the effect of different carbon dioxide concentrations in the future temperature changes, and the GFDL—CM3 method of sea surface temperature change is forecasted, found the red sea will experience a significant increasing trend [24]. Long (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 in local SST, so as to predict the trend of SST change from 2010 to 2040 [25]. Folland (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 in 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].

In terms of research content, 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 how the two factors cause and affect each other [29].

In the aspect of prediction methods, there are relatively a few studies on the prediction of SST variation. Most of them use neural networks, classical climate models or mathematical statistical methods. However, whether these methods are suitable for long-term SST prediction needs to be discussed. As for the most popular neural network methods, the model must be modified according to the data characteristics [30]. 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 [31]. 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 [32]. 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 [33]. And that is 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 [34]. 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 in the prediction time range. What's more, 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 [35].

Except for deep learning, 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]. Besides, few studies have explored the relationship between SST variation from the perspective of global warming, ignoring the effect of climate, one of the most important factors, on SST variation.

3 Preliminary work and proposed methods

These problems above are the characters of SST prediction, and this paper attempts to build a model with the ability to predict regional long-term sea temperature change [37]. The model can have a high prediction accuracy in a long-time span and investigate the impact of global warming on SST variation, which provides data support for improving people's understanding of sea surface temperature variation 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 variation is very complex, and it is difficult to quantify the influence of different factors on SST. Due to the complexity of SST variation, its changing progress is a Grey system on the whole. This paper hopes to propose a method to reveal the characteristics and degree of SST variation and predict the SST in the future based on the grey model.

Based on the analysis, the construction method of the model is described as follows:

Firstly, the steps of introducing the Grey model and getting the prediction result sequence by using the regional historical SST are introduced.

Moreover, a quantitative model of the impact of different greenhouse gas emission scenarios on future SST variation 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.

3.1 Grey model

Grey model (GM) 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 indexes data are incomplete or uncertain, the system with these characters is a Grey system. The Grey model is insensitive to data regularity, information content and information integrity [38]. After processing, the randomness of the system is significantly reduced, becoming a regular system with generation numbers. Therefore, the dynamic model in the form of differential equation is constructed, which has approximation and non-uniqueness.

For the prediction of SST, the Grey model helps us analyze the correlation between historical sea surface temperature observation data and infer the future seawater temperature trend through the dynamic change. After the existing temperature data is quantified, the generated number with strong regularity is obtained by the cumulative generation method, and then the model is obtained through the function change.

3.2 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, 40]. CMIP5 introduces a new greenhouse gas emission scenario, named representative concentration pathways (RCPs). 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 variation, 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, that is, unregulated emissions of carbon dioxide; RCP2.6 represents the minimum level of carbon dioxide emissions. 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.

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

The following conclusions can be drawn from the study of SST variation: SST changes are related to past sea surface temperatures, the 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 Model, 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. The different degrees of global warming in the future will also affect the degree and speed of SST variation.

Based on the above method, a Grey model of regional sea surface temperature under atmospheric reflection is established according to the characteristics of SST, named a Regional Sea Surface Temperature-Atmospheric Radiation Grey Model (RSST-ARGM).

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

4.1 Data checking and preprocessing

According to the requirements of RSST-ARGM, and the characteristics of data set in the Grey model, the data of SST needs to be to be preprocessed, in which the defects of the data set itself should be properly cleaned to ensure its integrity and effectiveness.

Because the SST data are collected from observed historical records, and the main 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, RSST-ARGM stipulated that if the missing rate of the data set is less than 5%, the data will be deleted

(the data of the month or year will not be counted, to prevent the artificial influence. If the missing rate is greater than or equal to 5%, the lost datum will be filled by taking the mean value of the two data before and after the missing datum, to reduce the introduction of new errors as much as possible.

The change of sea surface temperature is relatively flat, and the temperature in a region does not change dramatically.

Therefore, the error statistics were checked by using box plot: the data of every month 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 data at the same latitude at the same time as a reference sequence

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

2. Calculate the series ratio of the sequence above

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

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

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

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

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

4.2 Grey model prediction of SST

In RSST-ARGM, we take the data as the original sequence (0), offering the original data and generate a new row (1) through accumulation changes, to waken the randomness of the original SST data and exposure its 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 of SST after data inspection and preprocessing meets the requirements of the RSST-ARGM system, and the data is put into the system through the following steps:

1. Select the original data sequence from the SST

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

Add up to generate a new sequence

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

among them, $a^{(1)}(s) = \sum_{i=1}^s a^{(0)}(i)$, $s = 1, 2, \dots, n$

- Calculate the mean number of the sequence

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

So

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

The differential equation is

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

The corresponding albino differential equation is

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

After calculation, the SST data sequence is transformed into the corresponding differential equation.

- Solve the above equation

Mark-related parameter

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

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

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

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

- Test the predicted sequence.

According to the model requirements, if $\varepsilon(s) < 0.2$, $\varepsilon(s) = [a^{(0)}(s) - \hat{a}^{(0)}(s)]/a^{(0)}(s)$, $s = 1, 2, \dots, n$, $\hat{a}^{(0)}(1) = a^{(0)}(1)$, 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 Eq. (6) to calculate $\rho(s) = 1 - [(1 - 0.5x)/(1 + 0.5x)] \times \lambda(s)$. If $\rho(s) < 0.2$, it meets the requirement of RSST-ARGM. If neither $\varepsilon(s)$ nor $\rho(s)$ meet the requirements, we are required to re-examine the SST data set and readjust the model parameters until the results meet inspection standards.

4.3 RCPs

The temperature changes under different RCPs indexes are very other. Try to quantify the effect of different RCPs indices on temperature variation, we need to divide the

existing data set into prediction data set and detection data set. The influence factors of RCP 2.6, 4.5, 6.0 and 8.5 are defined as:

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

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:

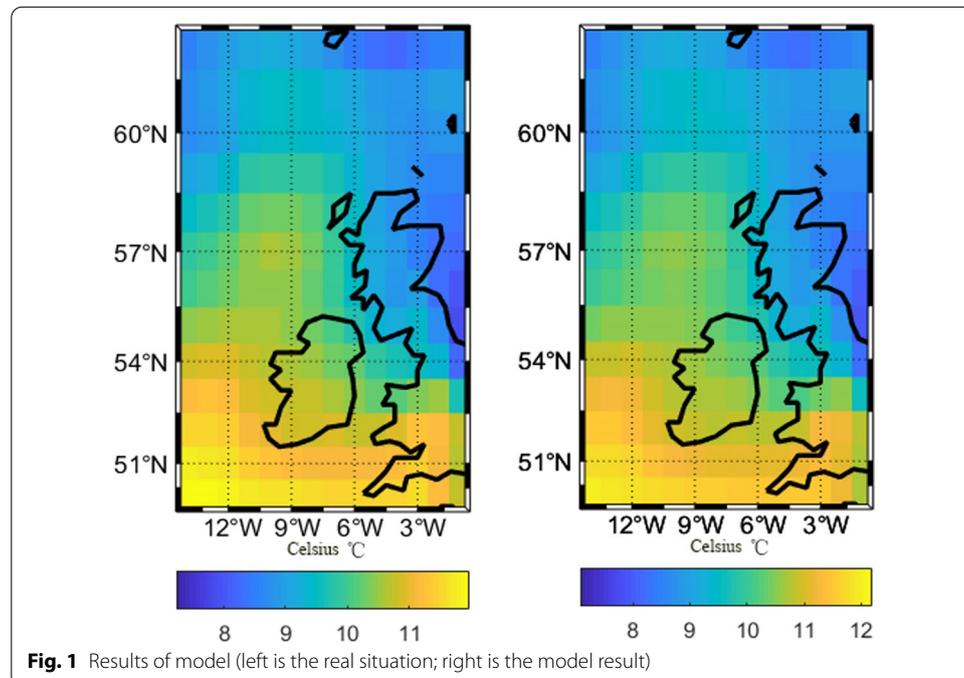
$$T(k + 1) = \gamma_i \times \hat{x}^{(1)}(k + 1), \quad k = 1, 2, \dots, n - 1, \dots$$

5 Result

Theoretical derivation alone is not enough. The simulation experiment is needed to prove that the RSST-ARGM prediction model has a high prediction accuracy. 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 shown in Fig. 1, explore local sea surface temperatures over the next 50 years.

5.1 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 data sets. 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.

5.2 Simulation design

The simulation experiment is divided into two parts. First, the accuracy of the prediction results of the model needs to be tested using historical data. If the accuracy meets the requirements, we use RSST-ARGM to predict future SST.

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

In order to improve the test efficiency, we mainly investigate the 2 months when the sea temperature is the highest (July and August). In the detection part, we chose the data from 1870 to 1966 to predict the results in 2016. In the prediction section, data from 1870 to 2020 are used to predict the SST over the next 50 years, from 2021 to 2070.

5.3 Simulation result

Because of the large amount of data, it is obviously difficult to test each data result. Therefore, in this paper, we need to analyze the prediction effect in the 50-year time range.

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 10.5° E, and 45.5° N to 65.5° N) successively, as shown in Fig. 1. Cosine similarity algorithm is used to compare the differences between two pictures [46]. According to the theory of cosine similarity algorithm, if the error cosine value is close to 1, the error angle is relative to 0° (the coincidence state), and the two vectors have high similarity. Conversely, the more error cosine value is close to -1, the more significant the difference.

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:

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

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

The test results show that the error cosine value of the sea temperature map based on the model's predicted sea temperature and the actual temperature is 0.99984, and

the error angle is 1.0115° at the resolution of 1 longitude times 1 latitude. Prove the validity of data prediction.

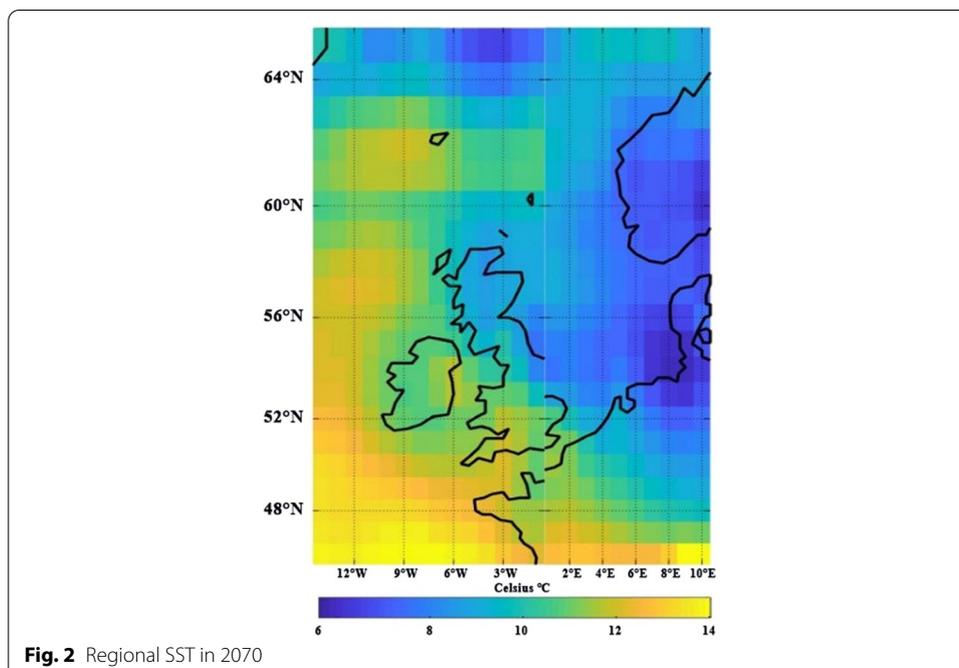
The 2016 comparison showed the accuracy of the model's predictions. RSST-ATGM model is used to predict the regional SST for the next 50 years. Based on the RSST-ARGM prediction model, we obtained sea surface temperature pictures 50 years later (2070) in the situation of RCP6.0, as shown in Fig. 2.

It can be seen from the model results that the July–August sea surface 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°C , which is 1.7°C 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 prediction result, 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 marine resources. Still, in the long term, this will no doubt cause more 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.



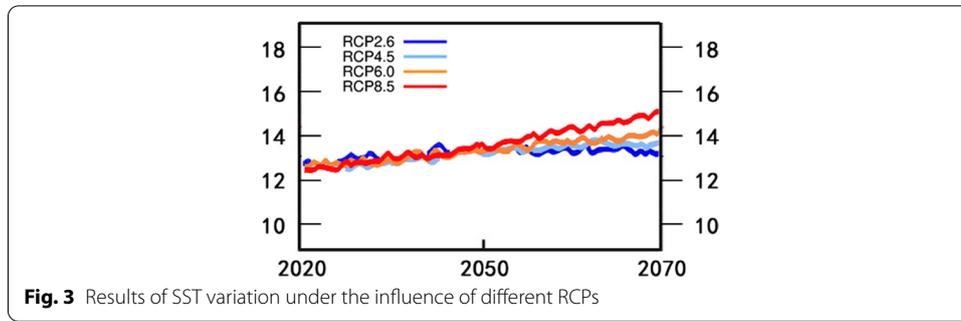


Table 1 Prediction result in 2017 under different RCPs

RCPs	Temperature (°C)	Trend
2.6	13.2	Stable
4.5	13.8	Rise slowly
6.0	14.2	Rise
8.5	15.3	Rise sharply

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 indexes is, SST keeps growing during 2021–2030, and the temperature of growth is similar to that of about 13 °C. After 2030, the seawater temperature at RCP2.6 tends to be stable, and will remain at 13.2 °C in 2070. RCP4.5, RCP6.0 and RCP8.5 increased continuously, reaching 13.8 °C, 14.2 °C and 15.3 °C respectively in 2070, as shown in Table 1. The comparison of temperature changing trend is shown in Fig. 3.

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

It is worth noting that even the predicted SST under RCP2.6 does not show an apparent downward trend after rising to 13.2 °C, 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 °C, the damage to the global climate is unimaginable.

6 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.

To make the comparison test representative, we choose the most classic BP neural network to compare with the RSST-ARGM. Using the same data set as model input, the prediction accuracy and time complexity of the two models were compared

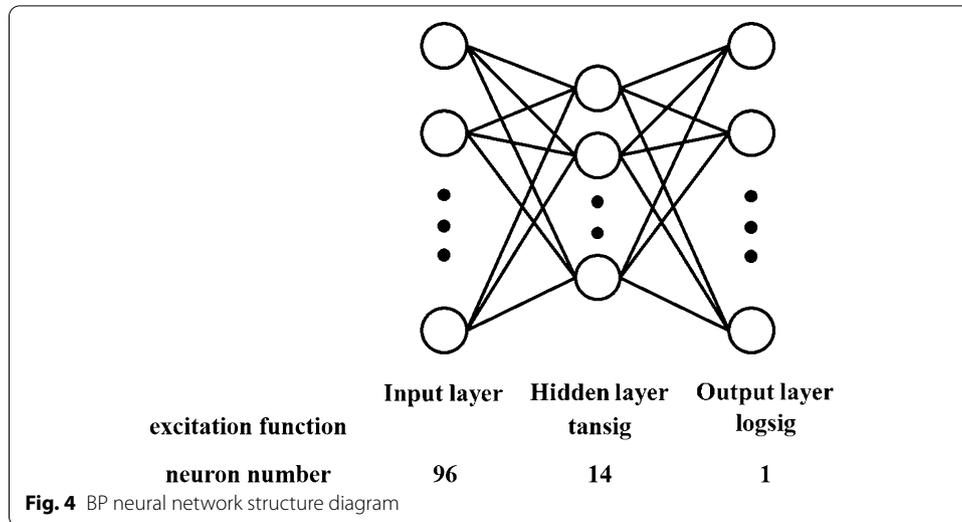


Table 2 Comparison of the prediction effects of the two models

	Training set	Prediction set	Angle (°)	Vector	Calculation time (min)
RSST-ARGM	1870–1966	2016	1.0115	0.99984	38.56
BP network			2.2445	0.99923	8.62

during the calculation process. We still use 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 set as the network training function, and the network iterations number matters a lot setting as 5000. To get good results, the expected error is 0.01 [49], as 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 Table 2.

The comparison in Table 2 directly shows the characteristics of the two models:

1. *Calculation time* Under the condition of the same prediction accuracy, BP network model needs a large amount of data and repeated training, which seriously affects the prediction efficiency of BP network, resulting in more than 80% of the calculation time of BP network is used for model training. On the other hand, RSST-ARGM model only needs to calculate the historical data set to get the predicted results without repeated training process, so it has a significant advantage in time complexity, and this advantage will become more obvious with the increase in data amount.
2. *Model applicability* In principle, there are some differences between the two models in their wide applicability. The prediction accuracy of BP network is based on model training, so the premise of using BP network model is to train a lot in specific areas, which means that the prediction scenarios of the model cannot be changed at will. However, the RSST-ARGM model only calculates the relationship between the data to get the prediction result, so it can make the temperature prediction for any water at any time.
3. *Prediction accuracy* Even though BP network model has carried out a lot of training, RSST-ARGM model is still better in the prediction accuracy using the same data set. The cosine similarity test compares the errors between the data predicted by the model and the actual data. The closer the error angle between the two is to 0, the higher the similarity between the two is. The error angle of RSST-ARGM prediction is only 1.0115° , while the error angle of BP network prediction is as high as 2.2445° . This fully indicates that the RSST-ARGM model is more suitable for the prediction of SST variation.

Through the comparison of three main features, 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 indexes 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 variation of SST.

7 Conclusion

In this paper, we combine the Grey model and representative concentration pathways in climatology, establishing a Regional Sea Surface Temperature-Atmospheric Radiation Grey Model, RSST-ARGM. Based on the calculation of historical SST, the model explores the trend of SST variation in the future, and introduces the influence of atmospheric greenhouse gas concentration change on SST variation in the form of RCPs indexes, to reflect the correlation between global warming and SST variation. Through simulation, in this article, the RSST-ARGM model is applied to predict the temperature variation 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 indexes has 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. In the future, the influence of different data set richness on prediction accuracy and prediction time length can be further investigated. Since the trend of SST variation 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

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 Intergovernmental Panel on Climate Change; CMIP3/5: Coupled Model Intercomparison Project; GM: Grey Model; ANN: Artificial Neural Network; CNN: Convolutional Neural Network; RSST-ARGM: A Regional Sea Surface Temperature-Atmospheric Radiation Grey Model; HadISST: Hadley Centre Sea Ice and Sea Surface Temperature.

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' contributions

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.

Authors' information

Linqian Zhu 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.

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 (Nos. 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).

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)

Declaration

Competing interests

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

Author details

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

Received: 20 April 2021 Accepted: 13 August 2021

Published online: 26 August 2021

References

1. M.-L. Timmermans, J. Toole, R. Krishfield, Warming of the interior arctic ocean linked to sea ice losses at the basin margins. *Sci. Adv.* **4**(8), 6773 (2018)
2. G. Zheng, X. Li, R.-H. Zhang, B. Liu, Purely satellite data-driven deep learning forecast of complicated tropical instability waves. *Sci. Adv.* **6**(29), 1482 (2020)
3. J.V. Ratnam, H.A. Dijkstra, S.K. Behera, A machine learning based prediction system for the Indian Ocean dipole. *Sci. Rep.* **10**(1), 284 (2020)
4. G. Hervieux, M.A. Alexander, C.A. Stock et al., More reliable coastal SST forecasts from the North American multimodel ensemble. *Clim. Dyn.* **53**, 1–16 (2017)
5. S. Qian, J. Chen, X. Li et al., Seasonal rainfall forecasting for the Yangtze River basin using statistical and dynamical models. *Int. J. Climatol.* **40**(1), 361–377 (2020)
6. D.F. Dias, A. Subramanian, L. Zanna et al., Remote and local influences in forecasting Pacific SST: a linear inverse model and a multimodel ensemble study. *Clim. Dyn.* **52**, 3183–3201 (2019)
7. S.J. Sohn, C.Y. Tam, H.I. Jeong, How do the strength and type of ENSO affect SST predictability in coupled models. *Sci. Rep.* **6**, 33790 (2016)
8. A. Capotondi, P.D. Sardeshmukh, E.D. Lorenz et al., Predictability of US West Coast ocean temperatures is not solely due to ENSO. *Sci. Rep.* **9**(1), 10993 (2019)
9. M. Ionita, V. Nagavciuc, Forecasting low flow conditions months in advance through teleconnection patterns, with a special focus on summer 2018. *Sci. Rep.* **10**, 1–12 (2020)
10. M. Taye, E. Dyer, K.J. Charies et al., Potential predictability of the Ethiopian summer rains: understanding local variations and their implications for water management decisions. *Sci. Total Environ.* **755**, 142604 (2020)
11. F. Counillon, N. Keenlyside, T. Toniazzo et al., Relating model bias and prediction skill in the equatorial Atlantic. *Clim. Dyn.* **775**, 1–14 (2021)
12. S. Kale, Development of an adaptive neuro-fuzzy inference system (ANFIS) model to predict sea surface temperature (SST). *Oceanol. Hydrobiol. Stud.* **49**(4), 354–373 (2020)
13. D. Hotta, Y. Ota, Statistical generation of SST perturbations with spatio-temporally coherent growing patterns. *Q. J. R. Meteorol. Soc.* **145**, 1660–1673 (2019)
14. M.G. Jacox, M.A. Alexander, C.A. Stock et al., On the skill of seasonal sea surface temperature forecasts in the California Current System and its connection to ENSO variability. *Clim. Dyn.* **53**, 7519–7533 (2017)
15. B. Narapusetty, The role of atmospheric internal variability on the prediction skill of interannual North Pacific sea-surface temperatures. *Theor. Appl. Climatol.* **133**(11), 1–9 (2017)
16. M. Newman, P.D. Sardeshmukh, Are we near the predictability limit of tropical Indo-Pacific sea surface temperatures? *Geophys. Res. Lett.* **44**, 8520–8529 (2017)
17. R. Aurélien, S. Qasmi, N.P. Gillett, Making climate projections conditional on historical observations. *Sci. Adv.* **7**(4), 0671 (2021)
18. R. Francisco et al., Climate impacts on global hot spots of marine biodiversity. *Sci. Adv.* **3**, 1601198 (2017)
19. Y. Ham, J. Kim, J. Luo, Deep learning for multi-year ENSO forecasts. *Nature.* **573**, 568–572 (2019)
20. F. Sévellec, S.S. Drifhout, A novel probabilistic forecast system predicting anomalously warm 2018–2022 reinforcing the long-term global warming trend. *Nature Commun.* **9**, 3024 (2018)
21. Y. Chikamoto, S.-Y. Wang, M. Yost et al., Colorado river water supply is predictable on multi-year timescales owing to long-term ocean memory. *Nature Rev. Earth Environ.* **1**, 26 (2020)
22. L. Hermanson, H.L. Ren, M. Vellinga et al., Different types of drifts in two seasonal forecast systems and their dependence on ENSO. *Clim. Dyn.* **51**, 1411–1426 (2017)
23. R. Seager, M. Cane, N. Henderson, D.-E. Lee et al., Strengthening tropical Pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nature Clim. Change.* **9**, 517–522 (2019)
24. M. Shaltout, Recent sea surface temperature trends and future scenarios for the red sea. *Oceanologia.* **61**, 484–504 (2019)
25. Z. Long, W. Perrie, Changes in ocean temperature in the Barents Sea in the 21st century. *J. Clim.* **30**(15), 5901–5921 (2017)
26. C.K. Folland, O. Boucher, A. Colman, D.E. Parker, Causes of irregularities in trends of global mean surface temperature since the late 19th century. *Sci. Adv.* **4**(6), 5297 (2018)
27. G. Melissa, S. Jeffrey, K. Yochanan, Mechanisms governing the development of the North Atlantic warming hole in the CESM-LE future climate simulations. *J. Clim.* **31**, 17–06351 (2018)
28. M. Ogurtsov, M. Lindholm, R. Jalkanen et al., North Atlantic sea surface temperature, solar activity and the climate of Northern Fennoscandia. *Adv. Space Res.* **59**(4), 980–986 (2016)
29. W. Linbo, Z. Hui, F. Xianyong, W. Shaohua, D. Weiping, G. Yanwen, A unified two-parallel-branch deep neural network for joint gland contour and segmentation learning. *Future Gener. Comput. Syst.* **100**, 316–324 (2019)
30. D. Songtao, Q. Shiru, X. Yuling, W. Shaohua, Stimulus-driven and concept-driven analysis for image caption generation. *Neurocomputing.* **398**, 520–530 (2020)

31. C. Chen, L. Bin, W. Shaohua, Q. Peng et al., An edge traffic flow detection scheme based on deep learning in an intelligent transportation system. *IEEE Trans. Intell. Transp. Syst.* **PP**(99), 1–13 (2020)
32. Y. Zhao, H. Li, S. Wan, S. Anjany et al., Knowledge-aided convolutional neural network for small organ segmentation. *IEEE J. Biomed. Health Inform.* **99**, 1–1 (2019)
33. Z. Gao, Y. Li, S. Wan, Exploring deep learning for view-based 3D model retrieval. *ACM Trans. Multimed. Comput. Commun. Appl.* **16**(1), 1–21 (2020)
34. S. Kiyohara, M. Tsubaki, T. Mizoguchi, Learning excited states from ground states by using an artificial neural network. *npj Comput. Mater.* **6**, 68 (2020)
35. S. Wan, X. Xu, T. Wang, Z. Gu, An intelligent video analysis method for abnormal event detection in intelligent transportation systems. *IEEE Trans. Intell. Transp. Syst.* **22**(7), 4487–4495 (2021)
36. N. Cuvelier, S.M. Bartell, Shrinkage estimation of long-term water ingestion rates. *J. Exposure Sci. Environ. Epidemiol.* 1–9 (2021)
37. P.J. Athanasiadis, S. Yeager, Y.-O. Kwon, D.W. Bellucci, A. Smith et al., Decadal predictability of North Atlantic blocking and the NAO. *npj Clim. Atmos. Sci.* **3**, 20 (2020)
38. W. Edeling, H. Arabnejad, R. Sinclair, D. Suleimenova et al., The impact of uncertainty on predictions of the CovidSim epidemiological code. *Nature Comput. Sci.* **1**(2), 128–135 (2021)
39. P. van Vuuren Detlef, E. Jae, K. Mikiko et al., The representative concentration pathways: an overview. *Clim. Change.* **109**, 5 (2011)
40. M.G. Bréon, O. Boucher, F.M. Bréon et al., Declining uncertainty in transient climate response as CO2 forcing dominates future climate change. *Nature Geosci.* **8**(3), 181–185 (2018)
41. R. Ganzemüller, P. Pradhan, J.P. Kropp, Sectoral performance analysis of national greenhouse gas emission inventories by means of neural networks. *Sci. Total Environ.* **656**, 80–89 (2019)
42. X. Han, J. Chang, A hybrid prediction model based on improved multivariable grey model for long-term electricity consumption. *Electr. Eng.* **103**(1), 1031–1043 (2020)
43. H. Wang, J. Zhang, W. Jin, Z. Yan, in *Research on mangrove recognition based on hyperspectral unmixing*. Proceedings of 2017 IEEE International Conference on Unmanned Systems (ICUS), vol. 10 (2017), pp. 312–314
44. J. Lee, W.W. Wasserman, G.F. Hoffmann, C. Karnebeek et al., Knowledge base and mini-expert platform for the diagnosis of inborn errors of metabolism. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **20**(1), 151–158 (2018)
45. Z. Hausfather, K. Cowtan, D. Clarke, P. Jacobs et al., Assessing recent warming using instrumentally homogeneous sea surface temperature records. *Sci. Adv.* **3**(1), 1601207 (2017)
46. W. Xiang, Y. Li, R. He, M. Gao et al., A novel artificial bee colony algorithm based on the cosine similarity. *Comput. Ind. Eng.* **115**(JAN.), 54–68 (2017)
47. H. Zhang, X. Lu, in *The prediction of PM2.5 value based on ARMA and improved BP neural network model*. 2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS), vol. 81 (2016), pp. 515–517
48. S. Tamura, Y. Nishitani, C. Hosokawa, T. Miyoshi et al., in *Multiplex communication by BP learning in neural network*. 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), vol. 10 (2016), pp. 825–828
49. Z. Yini, Z. Ping, Y. Zhongjiang, M. Yang et al., in *The research for a kind of information fusion model based on BP neural network with multi position sources and big data selection*. 2019 IEEE 2nd International Conference on Electronics Technology (ICET), vol. 5 (2019), pp. 619–623

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
