https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

An interpretable machine learning for marine heatwave prediction for the south China sea

Peihao Yang[1], Guodong Ye[1,2,*]

[1]College of Ocean and Meteorology, Guangdong Ocean University, Zhanjiang 524088, China

[2]Faculty of Mathematics and Computer Science, Guangdong Ocean University, Zhanjiang 524088, China

***Corresponding Author:** Guodong Ye (guodongye@hotmail.com)

Abstract

A primary challenge of machine learning to predict marine heatwave (MHW) for the south China sea (SCS) is the limited availability of observational data for model training. To address this issue, this study explores the viability of leveraging multi-member ensemble simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6), to construct an extensive, physically consistent training dataset for various machine learning models. After training on multiple CMIP6 ensemble members, the constructed models are evaluated for their predictive capacity regarding MHW in the SCS. The results also show that these machine learning-based methods can perform comparably to the existing dynamic models in terms of prediction performance, and in some cases even outperform the latter. Furthermore, by incorporating machine learning interpretability techniques, the key physical processes can also be elucidated from these predictions. That is to say, the new method is not a traditional "black box", but rather an effective tool that can possess certain physical transparency and scientific interpretability.

Keywords: South China Sea marine heatwave; machine learning; interpretable analysis; marine heatwave prediction.

1 Introduction

Marine heatwaves (MHWs) are extreme oceanic climate events characterized by abnormally elevated sea surface temperatures, persisting from days to several months and even extending across thousands of kilometers (Hobday et al., 2016; Hobday et al., 2018). In recent years, the frequency of MHWs has risen markedly under global warming (Frölicher et al., 2018; Oliver et al., 2018; Holbrook et al., 2020). These abnormal high-temperature events severely impact marine ecosystems, associated with ecosystem services, and the economy (Perkins-Kirkpatrick et al., 2019; Smale et al., 2019). The 2014-2016 "Blob event" in the North Pacific caused a large-scale disruption to marine ecosystems, resulting in mass marine mortality and severe declines in fishery resources (Di Lorenzo et al., 2016). China has a vast offshore area, among which the SCS in the tropical zone experienced frequently many MHWs (Yao et al., 2020; Xiao et al., 2019). The SCS has coral reef habitats and modern marine pastures, and MHWs in the SCS caused severe impacts on its marine ecosystem and economy (Feng et al., 2022; Mo et al., 2022; Zhao et al., 2023). Considering the significant impacts of MHWs on the ecological environment, fishery production, and economic activities, enhancing the predictive capability of MHWs has become a top priority in marine resource management.

MHWs are modulated by local and large-scale air-sea interactions (Lee et al., 2010; Chen et al., 2015; Tang et al., 2025). Local atmospheric processes are generally closely related to the persistent abnormal high-pressure system in the upper atmosphere. Specifically, this may manifest as reduced cloud cover, enhanced short-wave radiation, and weakened wind speed, which leads to reduced evaporation (Amaya et al., 2020; Sen Gupta et al., 2020). Local oceanic processes include anomalous horizontal and vertical heat advection, weakened vertical mixing, and a shoaling of the mixed layer (Amaya et al., 2021; Han et al., 2022). MHWs are also influenced by remote forcing or oceanic modulation associated with large-scale climate models such as ENSO, the Indian Ocean Dipole (IOD), and the North Atlantic Oscillation (NAO), though the dominant mechanisms may differ substantially across regions (Qi et al., 2022; Saranya et al., 2022). For instance, the local climate of the SCS is influenced by the upper-level ocean currents, and these currents are mainly regulated by the East Asian monsoon (Gan et al., 2006). Local abnormal anticyclones enhance shortwave radiation, weaken the southwest monsoon and upwelling, while abnormal marine anticyclones exacerbate water convergence and Ekman downwelling, further intensifying sea surface warming and increasing the probability of MHWs in the SCS (Yao et al., 2021; Liu et al., 2022; Tan et al., 2022). These anomalous anticyclonic processes may be modulated by teleconnections from large-scale climate models. Therefore, predictive models can improve the stability and reliability of MHW forecasting by capturing these potential predictability signals.

MHW prediction methods can be broadly categorized into dynamic methods, empirical methods (statistical techniques or machine learning techniques), and hybrid methods (integrating dynamic techniques with empirical techniques). Dynamical prediction methods typically rely on ensemble forecasting, which performs probabilistic prediction through the distribution of ensemble members.

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

The prediction uncertainty originates from minor errors in initial conditions, which grow rapidly over time (Tao et al., 2017; Waliser et al., 2003). By calculating the mean of ensemble members, random noise from individual weather events can be filtered out to extract more stable climatic signals. Statistical prediction methods have a long history, among which Canonical Correlation Analysis (CCA) is one of the classic methods (Barnston et al., 1996; Rana et al., 2018). The CCA is typically used to describe the linear relationship between two sets of variables. It is widely used in the relationships between sea surface temperature and precipitation or atmospheric circulation fields to identify possible teleconnection patterns. However, the traditional CCA framework has limitations in handling multiple predictor variables and their interactions simultaneously. Particularly in the case of high-dimensional data or a large number of predictor variables, overfitting is prone to occur (Chen et al., 2017). Moreover, the CCA has insufficient capability to characterize temporal evolution, and can only make up for the lack of temporal information to a certain extent through lagged correlation.

Currently, significant research resources have been invested in dynamic model ensemble forecasting, but studies on applying machine learning to MHW prediction remain relatively scarce (Slater et al., 2023). A major challenge in machine learning for MHW prediction lies in the limited availability of observational data for model training (Gao et al., 2024). This limitation not only hinders traditional statistical methods but also affects machine learning-based approaches. For example, to reliably simulate the nonlinear interactions between multiple predictor variables, each predictor variable requires a large number of samples to avoid model overfitting (Van der Ploeg et al., 2014). However, the available observational samples typically span only around four decades, which is insufficient to support large-scale machine learning training. An effective alternative is to train machine learning models on large ensembles of climate model simulations (Seferian et al., 2020; Li et al., 2021). This approach substantially expands the training dataset, enabling coverage of multiple centuries of simulated climate conditions. For example, a study has found that training a Convolutional Neural Network (CNN) on historical climate model simulation data can achieve high-precision ENSO prediction with a lead time of over one year (Ham et al., 2019). In addition, some studies used regularized models to achieve high-precision predictions with various meteorological datasets (Jeffree et al., 2024; Kim et al., 2017).

Due to the massive data generated by increasing satellite observations and climate models, machine learning algorithms demonstrate enormous potential in early warning of upcoming extreme events. This study focuses on five machine learning models: Random Forest, XGBoost, LSTM, MLP and SVR. Because they show excellent performance in recent studies for predicting climatic variables such as precipitation and sea surface temperature (Park et al., 2016; Ham et al., 2023; Sattari et al., 2021), as well as for predicting the impacts of climatic extremes (Zhu et al., 2021; Lin et al., 2020; Yang et al., 2025). The objective is to evaluate the feasibility of using large-ensemble climate model data and various machine learning techniques to predict MHWs in the SCS. Then,

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

machine learning models are constructed and applied to predict MHWs based on several climatic variables. The models are trained and validated using publicly available climate model samples, observational samples, and reanalysis datasets to assess the predictability of oceanic extremes across various forecast lead times. Furthermore, interpretable machine learning methods are employed to identify the key physical processes affecting predictions, providing physical interpretability support for the forecast results.

2 Related knowledge

2.1 LSTM

Long Short-Term Memory (LSTM) is a special type of recurrent neural network (RNN). It can effectively capture long-term dependencies in time series data (Yu et al., 2019). Traditional RNNs often suffer from gradient vanishing or gradient explosion when handling long sequences. This makes it difficult for them to learn long-range dependencies. LSTM addresses this problem by introducing memory cells and gating mechanisms. These gates include an input gate, a forget gate, and an output gate. The forget gate decides which information should be discarded. The input gate controls how new information is stored. The output gate determines the final output. These mechanisms allow LSTM to pass information effectively across different time steps. These include convolutional neural networks (CNNs) and attention mechanisms. This helps improve its ability to model complex temporal data. In this study, LSTM model consists of three unidirectional layers. Each layer has 128 hidden units. The model is trained with a batch size of 64 and a learning rate of 0.001. Loss function is set to be mean squared error (MSE). Training is performed on standardized sliding-window data for 100 epochs. Dropout regularization is applied during training to reduce the risk of overfitting.

2.2 MLP

Multi-Layer Perceptron (MLP) is a classical feedforward neural network. It consists of an input layer, one or more hidden layers, and an output layer (Cabaneros et al., 2019). Each neuron is fully connected to all neurons in the previous layer. It uses weights, biases, and activation functions such as ReLU, Sigmoid, or Tanh to perform nonlinear transformations. These operations allow the network to learn complex mappings. The core computations of an MLP include forward propagation and backpropagation. Backpropagation uses gradient descent to optimize the loss function and adjust network parameters to reduce errors. MLP has strong nonlinear modeling capability. Theoretically, it can approximate any continuous function. In this study, the MLP model has two hidden layers. Each layer contains 64 neurons. Activation function and loss function are set to be ReLU and MSE respectively. The model is trained for 100 epochs using features expanded through a sliding window. The optimizer is Adam with a learning rate of 0.001. Dropout regularization is

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

applied during training to reduce overfitting.

2.3 RF

Random Forest (RF) is a decision tree method using ensemble learning. It can improve the generalization ability by training multiple decision trees and using voting or averaging for prediction (Hu et al., 2023). The core idea of RF is to use the Bagging method. It draws multiple subsets from the original dataset through bootstrap sampling and trains a separate decision tree on each subset. During tree construction, each tree randomly selects a subset of features for splitting. This reduces the risk of overfitting and improves model stability. RF has strong robustness and can evaluate feature importance. It performs well when dealing with high-dimensional data and complex pattern recognition tasks. However, RF needs to train a large number of decision trees, which leads to high computational costs. On large datasets, it may face limitations in memory and computing resources. In this study, the RF model consists of 1,000 decision trees, each with a maximum depth of 10. Each leaf node contains at least 5 samples.

2.4 SVR

Support Vector Regression (SVR) is an extension of Support Vector Machines applied to regression tasks. It is used to find an optimal hyperplane that fits the data as closely as possible within a certain error margin (Roy et al., 2023). SVR uses kernel functions such as linear, radial basis function, and polynomial kernels to map data into a high-dimensional space. This enables it to model nonlinear relationships effectively. However, SVR has high computational complexity. It requires long training time and large memory usage when applied to large datasets. Its performance also depends on several hyperparameters such as C, ϵ, and the kernel parameters. These usually need to be tuned using cross-validation to ensure good prediction results. Therefore, SVR performs well in small-scale and high-precision regression tasks. But in large-scale applications, its efficiency and scalability may be limited. In this study, a radial basis function kernel is used. The regularization strength is set to C = 1.0. The kernel coefficient is automatically determined based on the feature dimensions.

2.5 XGB

XGBoost (Extreme Gradient Boosting) is an optimized algorithm using Gradient Boosted Decision Trees (GBDT). It improves predictive performance by iteratively building decision trees to minimize the loss function and adjust sample errors through weighted updates (Niazkar et al., 2024). XGBoost uses a variety of optimization strategies, such as second-order derivative updates, regularization, prevention of overfitting, parallel computation, and pruning. It can also handle missing values and automatically learn optimal split strategies, maintaining high computational efficiency even when applied to large-scale data. Compared with traditional GBDT, the optimization strategies of XGBoost significantly enhance training speed and generalization ability, making it highly effective for regression and classification tasks on structured data. In this study, the XGB

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

model uses 1,000 trees with a maximum depth of 10, a learning rate of 0.1, and a row and column sampling rate of 80 percent to improve predictive accuracy while controlling overfitting.

2.6 Feature Processing

This study applies Empirical Orthogonal Function (EOF) analysis (Martinez-Sosa et al., 2021) and lag correlation analysis (Zhang et al., 2022) to extract the coupled ocean-atmosphere spatial patterns and temporal characteristics associated with MHWs in the SCS. MHWs are usually influenced by large-scale ocean-atmosphere interactions, and spatial information from a single variable often contains redundancy. EOF analysis can effectively identify dominant modes of spatial variability. This dimensionality reduction helps highlight key climate variability features and reduces the impact of redundant information on machine learning model performance. Firstly, principal component analysis is conducted on the CMIP6 data and observational data using EOF to extract the main spatial modes of each variable within the study area (the South China Sea and its surrounding regions). Next, lag correlation analysis is taken to determine significant lead-lag relationships between climate variables and MHWs. This process aimed to capture precursor signals and provide effective predictive indicators for the model. Finally, the dominant spatial patterns obtained from EOF analysis and the key lead times identified through lag analysis are used as input features for the model to increase the accuracy and lead time of MHW prediction in the SCS.

3 Data and Method

3.1 Data

This study uses multi-model large ensemble simulation data from CMIP6 to provide sufficient training samples for the machine learning models. The historical simulations are conducted by several global climate models (GCMs) developed by different research institutions. The data cover the period from 1950 to 2014. These simulations follow historical forcing scenarios. They include observed changes in greenhouse gas concentrations, solar radiation, volcanic eruptions, and aerosols (Eyring et al., 2016). In this study, CMIP6 models include 20 high-resolution coupled climate models such as ACCESS-CM2, BCC-CSM2-MR, CMCC-CM2-SR5, MIROC6, and MRI-ESM2-0. These models have a spatial resolution of about 1 degree. They simulate the full interactions among the atmosphere, ocean, land and sea ice.

CMIP6 climate variables used in this study contain sea surface temperature, surface air temperature, sea level pressure, and zonal and meridional wind fields at 10 meters height. Previous studies show that CMIP6 models perform better than CMIP5 in simulating large-scale climate variability in the tropical and subtropical regions. They are especially better at reproducing the frequency, intensity, and teleconnection patterns of ENSO events (Bellenger et al., 2014). This is important for training machine learning models with climate model data. Systematic biases in teleconnections or air-sea interactions may be learned by the model during training and affect prediction results (Zhang et al., 2019).

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

After training the machine learning models using CMIP6 multi-model data, the offline trained models are applied to observational data for prediction testing. To ensure consistency of variables between the training and validation stages, same set of variables are used during testing. The sea surface temperature data came from the high-resolution Optimum Interpolation Sea Surface Temperature (OISST) dataset provided by NOAA. This dataset combines multiple satellite observations and in-situ measurements. It is commonly used in global climate studies and marine extreme event monitoring (Huang et al., 2017). Atmospheric circulation data were obtained from the ERA5 global reanalysis product provided by ECMWF. This dataset has high spatial and temporal resolution and includes a wide range of atmospheric variables. It was widely used in climate diagnostics and model evaluation (Hersbach et al., 2020). In processing the observational and reanalysis variables, the same dimensionality reduction and preprocessing methods as used for the CMIP6 data. This ensures consistency in the data processing and improved the generalization ability of the models under real observation conditions.

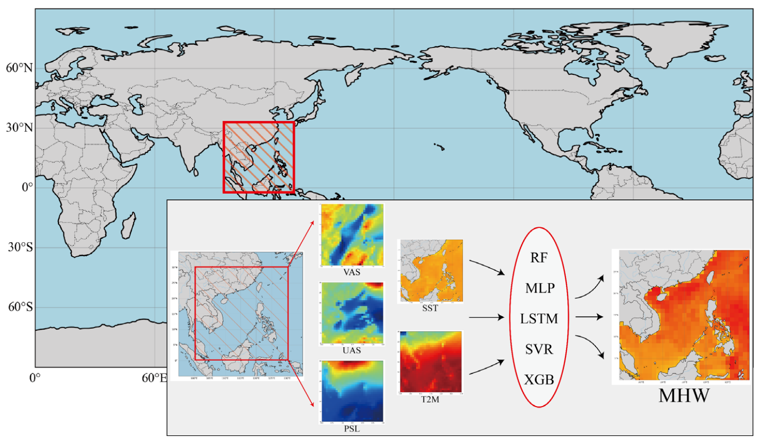

Fig.1. Overview of the marine heatwave prediction framework. In this study, the simulation experiments are conducted over the South China Sea, defined by the region spanning 0°-30°N and 100°-130°E. Five machine learning models are trained on CMIP6 ensemble data and subsequently tested using observational datasets to forecast MHWs in the SCS.

3.2 Method

The Fig. 1 shows the complete process of the framework. This study focuses on a specific region of the South China Sea, as shown by the red box in Fig. 1. SST from this region were extracted and used to construct a one-dimensional time series. To avoid the influence of long-term warming trends on model training, the SST series was first detrended. A daily climatological mean was then constructed to represent the typical SST for each day. Because high-frequency fluctuations exist in daily climate data, a 31-day moving average was applied to smooth the climatological mean and improve the stability of the analysis. SSTA was also computed by subtracting the climatological mean from the corresponding daily SST. Based on the SSTA, MHW events during the study period were identified and labeled for each sample. To reduce the impact of short-term natural fluctuations such as storms and intraseasonal disturbances, a 7-day moving average was applied. This helps improve the model's response to climate-scale signals and reduces the risk of overfitting. Other key meteorological drivers, including T2M, PSL, UAS, and VAS, were processed using the same method to ensure consistency across variables.

All input variables are standardized to improve comparability across different scales and to accelerate model convergence. EOF analysis and lag analysis are applied to the predictors, and the derived variables are used for model training. Five machine learning models are trained and calibrated using CMIP6 data. After training, all model parameters are kept fixed. The models are then applied offline to observational and reanalysis data to evaluate their performance in predicting MHWs in the SCS during the period 2015 to 2022. The processed predictors and labels are finally input into multiple machine learning models to perform classification and regression tasks for both short-term and long-term MHW prediction. Hyperparameters re tuned using a validation set to determine the optimal configuration. This workflow improves the stability and accuracy of the models and also laid the foundation for subsequent interpretability analysis using SHAP values and feature importance evaluation.

4 Results and Discussion

4.1 Prediction accuracy

After training and hyperparameter tuning of each machine learning model using a large ensemble data from CMIP6. In this study, model parameters are fixed after determination and observation data are used for multi-duration testing and evaluation. The prediction tasks span seven forecast lead times: 7, 15, 30, 90, 180, 365, and 730 days. In each prediction, the model uses observed atmospheric and oceanic variables prior to the target date as input features to simulate the prediction process in real-world application scenarios. This independent testing set evaluation provides an objective assessment of the model's generalization capability.

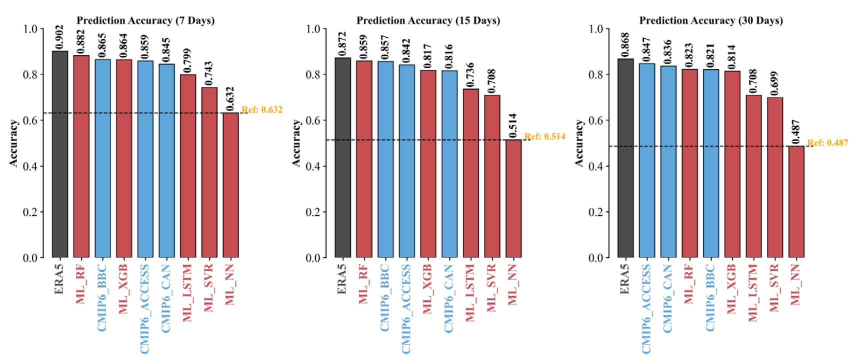

Fig.2. Comparison of prediction accuracy for MHWs in SCS across short-term lead times (7, 15, and 30 days) using three categories of methods. Black bars represent the prediction accuracy of ERA5 reanalysis data, blue bars denote the prediction accuracy of three CMIP6 model members (ACCESS, BBC, CAN), and red bars signify the prediction accuracy of five machine learning models (RF, XGB, LSTM, SVR, MLP). The values on each bar indicate the corresponding model's prediction accuracy.

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

To systematically compare the predictive capabilities of different models in various time spans, Fig. 2 presents the accuracy of five machine learning models for short-term predictions at 7, 15, and 30 days. The prediction accuracy of each machine learning method (red bars) is compared with the simulation results of three groups of CMIP6 members (blue bars) and the ERA5 reanalysis data results (black bars). All methods use the same test years to ensure horizontal comparability. Furthermore, Fig. 3 illustrates the performance of the same five machine learning models for longer-term forecasts at 90, 180, 365, and 730 days. Fig. 3 adopts the same comparison benchmark and color scheme as Fig. 2, facilitating analysis of the changing trends in model performance as prediction duration increases.

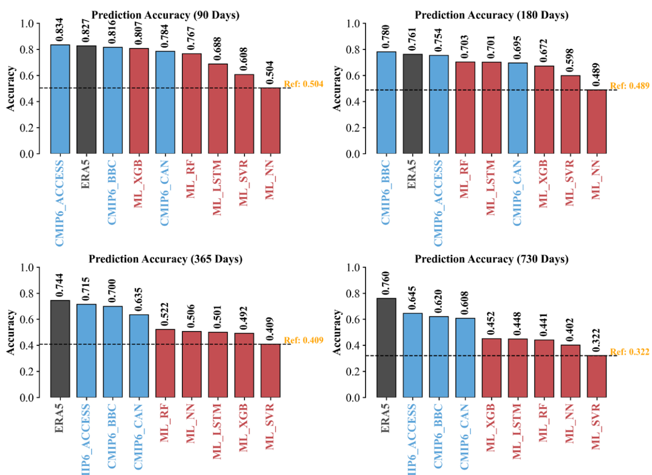

Fig.3. Comparison of prediction accuracy for MHWs in SCS across long-term lead times (90, 180, 365 and 720 days) using three categories of methods. Similarly to Figure 2, black bars represent the prediction accuracy of ERA5 data, blue bars denote the prediction accuracy of three CMIP6 model members, and red bars signify the prediction accuracy of five machine learning models. The values on the bars indicate the prediction accuracy, while the horizontal black dashed line represents the benchmark level for random predictions, with the benchmark decreasing gradually as the prediction duration extends.

In Fig. 2, ERA5 data, three groups of CMIP6 members (ACCESS, BBC, CAN), and five machine learning models (RF, XGB, LSTM, SVR, MLP) are compared for their accuracy in MHW prediction across three short-term prediction horizons (7 days, 15 days, and 30 days, respectively). From an overall perspective, it can be observed that the performance differences among models are relatively small over short prediction horizons, and the accuracy is generally high. ERA5 attains the highest prediction accuracy at the 7-day forecast horizon. The performance of the RF model closely follows that of the ERA5 data and is competitive with the ERA5 data. At 15- and 30-day lead times, RF model and XGB model consistently maintain leading performance. Notably, the RF model ranks among the top performers in the short-term and exhibits remarkable stability with the smallest precision fluctuations across different timesteps. It is also worthy of note that the ACCESS and BBC models have exhibited highly consistent performance in short-term predictions. Their performance

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

in multiple forecast horizons is almost comparable to that of machine learning models, and they even outperform certain machine learning models at specific time points (e.g., the 30-day). This demonstrates that physics-based model simulations retain a degree of advantage at specific lead times. This stability may be attributed to its long-term operational foundation and systematic modeling capability for physical processes. However, the MLP model and SVR model demonstrated the poorest performance in all forecast horizons, indicating their limited capability to extract complex signals from MHWs.

Fig. 3 illustrates performance of different models over longer timeframes (90 days, 180 days, 365 days, and 730 days). As the prediction lead time increases, the prediction accuracy of all models shows a decline to varying degrees, among which the decline in machine learning models is the most significant. the prediction accuracy of the RF model drops from 0.882 for 7-day to 0.522 for 365-day, and further drops to 0.441 for 730-day, representing a decline of over 40% in accuracy from 7 days to 730 days. The MLP model and SVR model exhibit more pronounced weakness in long-term predictions, indicating their unsuitability for identifying and forecasting MHW trends in extended periods. By contrast, ERA5 and CMIP6 demonstrate stronger stability, maintaining an accuracy above 0.6 over both 365-day and 730-day. This demonstrates the superiority of physical models in long-term climate prediction. Machine learning models perform outstandingly in short-term predictions, mainly benefiting from their non-linear fitting capability and high sensitivity to patterns in the training dataset. However, their predictive capability rapidly declines in long-term due to the absence of physical process constraints. Conversely, physical models such as CMIP6 and ERA5 exhibit stronger systematicity and stability, particularly demonstrating advantages in simulating teleconnection signals and low-frequency climatic backgrounds.

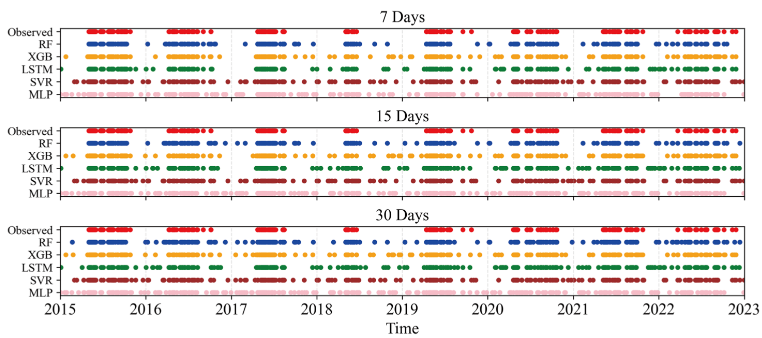

Fig.4. Comparison between prediction results of models in short-term and actual South China Sea MHWs. The scatter plot shows the occurrence of MHWs and the prediction results of different machine learning models. The scatter points in the first row represent actual MHWs, while subsequent rows denote predicted MHWs from each machine learning model.

To further evaluate the temporal detection capabilities of each model in different lead times, Fig. 4 and 5 present point-by-point comparisons between model predictions and observed MHWs. Each dot in the plots represents an individual MHW event, and each row corresponds to the prediction

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

output of a specific method. The top row shows actual MHWs identified from observational data, while the remaining rows present simulation results from different machine learning models in various prediction lead times. This point-by-point aligned visualization intuitively demonstrates the temporal fitting capability and accuracy of models in different time scales.

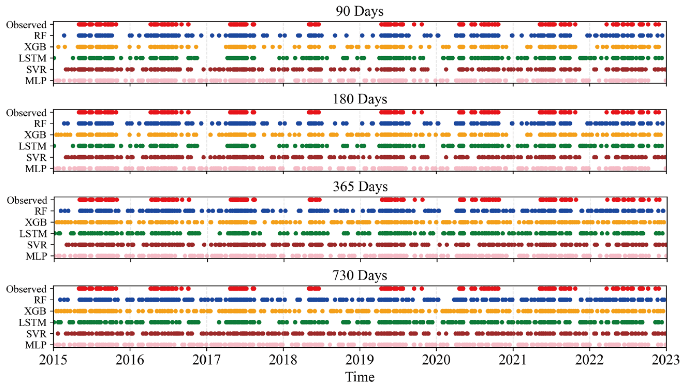

Fig.5. Comparison between prediction results of models in long-term and actual South China Sea MHWs. The scatter plot shows the occurrence of MHWs and the prediction results of different machine learning models. The scatter points in the first row represent actual MHWs, while subsequent rows denote predicted MHWs from each machine learning model.

Fig. 4 demonstrates the prediction performance of models in three short-term forecast lead times (7-day, 15-day, and 30-day). Most model predictions generally align with the distribution of actual MHWs, indicating that machine learning models can relatively accurately capture MHWs. Especially in the years when MHW occurred frequently such as 2016, 2020 and 2022, the alignment degree between the model and the observations was relatively high. This indicates that machine learning models exhibit good temporal localization capability about MHWs in short-term forecast lead times and can effectively respond to disturbances in short-term climatic signals. However, Fig. 5 shows a trend of gradual performance degradation in models as the forecast lead time increases from 90 to 730 days. It specifically manifests as the gradual dispersion of prediction result distributions from machine learning methods, featuring not only more deviating points but also substantial occurrences of "false positives" and "false negatives". In 365-day and 730-day predictions, the fitting capability of almost all models to actual MHWs significantly diminishes, with MHW prediction points overall deviating or dispersing, making it difficult to accurately reproduce the concentrated distribution intervals of events in observational data. For example, the performance of RF model and XGB model significantly declines in long-term prediction, while MLP model and SVR model exhibit near-random prediction states over extended durations. Overall, while machine learning models perform well in short-term predictions, their prediction accuracy declines in longer-term forecasts.

4.2 Interpretable Random Forest Model

Machine learning methods are often considered a relatively "brute-force" solution strategy, and

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

their "black-box" results in weak interpretability. The lack of interpretability in models may affect their credibility in scientific research and practical applications. Models sometimes produce superficially accurate results for erroneous mechanisms. Therefore, to further understand the internal operational mechanisms of the models, this section investigates the machine learning "black box" and attempts to reveal the physical implications in its decision-making process. RF model performs outstandingly in multiple prediction evaluations, and selects three representative forecast lead time—7-day, 30-day, and 90-day—for in-depth interpretation. It is important to emphasize that these interpretation methods merely serve as processing tools for model outputs and do not interfere with or affect the prediction pipeline or parameter architecture. The feature importance evaluation method in RF models relies on statistically averaging the improvement in classification purity contributed by each feature across all decision trees. At each node split, the "purity" reduction caused by the feature is calculated and multiplied by the number of samples at that node. The "impurity" reductions caused by the feature across all trees are summed up and normalized to obtain its importance score (Wang et al., 2024). This method can identify the most influential input features in the prediction process to provide a basis for the physical interpretability.

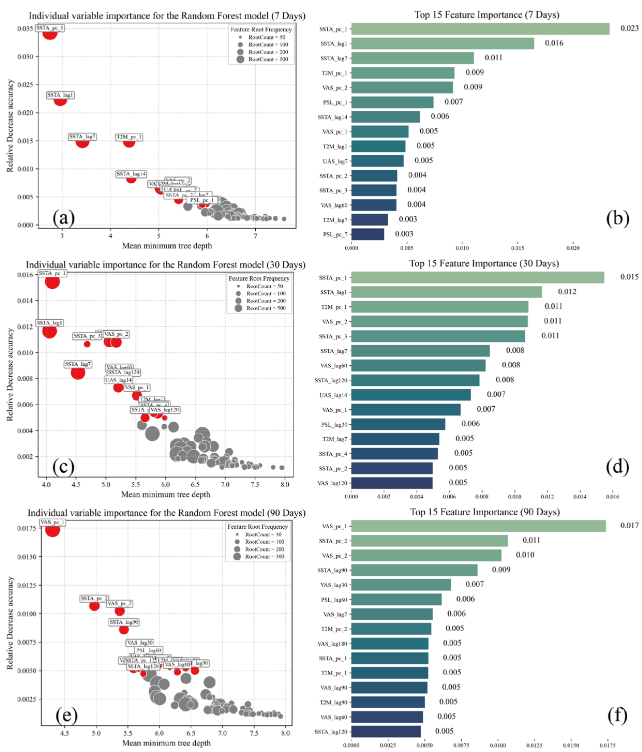

Fig.6. Univariate feature importance analysis in the Random Forest model. This figure presents univariate feature importance for Random Forest MHW forecasts at multiple lead times. It reports four indicators: feature importance scores from training that combine split frequency and information gain with higher values meaning greater contribution, relative accuracy reduction computed by shuffling a single variable with a larger drop implying more importance, average minimum tree depth showing where a variable first splits with shallower depth indicating higher

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

importance, and root node occurrence count reflecting how often a variable becomes the root with higher counts indicating stronger initial influence. Panels (a)(c)(e) mark the top 15 predictors in red matching the scores in (b)(d)(f), and names follow Fig. 1. For example, SSTA_pc_1 is the first EOF mode of SSTA and SSTA_Lag1 is a one-day lag.

Feature importance assessment is firstly employed to quantify the relative contribution of various predictors in the regression-based forecasting task. The results are presented in Fig. 6. Based on integrating multiple metrics, sea surface temperature anomaly (SSTA), 2-meter air temperature (T2M), and wind field information (UAS, VAS) emerge as the variables most significantly influencing prediction accuracy. Notably, there is a widespread negative correlation between the relative accuracy reduction metric of variables and their average minimum tree depth. This indicates that the most important variables tend to appear earlier in the upper layers of decision trees (closer to the root node). The dominance of key predictive factors also varies across different forecast lead times. SSTA and VAS consistently serve as dominant predictors in all lead times. In short-term forecasts (e.g., 7- and 30-day), SSTA-related features (including principal components and lagged terms) exhibit the highest importance, indicating they carry more information about recent MHWs. However, in longer-term forecasts (e.g., 90 days), the importance of VAS-related features increases significantly while the ranking of SSTA-related variables declines relatively, indicating that wind field signals may have greater predictive value on longer timescales. Additionally, T2M is relatively important in short-term forecasts while VAS plays a more critical role in long-term predictions. The trend of predictive variable weights changing with forecast lead times is also reflected in the lag period. The RF model tends to use variables with shorter lag periods for short-term forecasts, while input factors with longer lag periods gradually replace short-lag variables in long-term predictions. This change aligns with the expected pattern that variables closer to the target date generally contain more immediate information relevant to the prediction. It is worth emphasizing that the key predictive factors highlighted in Fig. 6 align with the current understand of formation mechanisms of MHWs in the SCS. Numerous studies have shown that large-scale climate models, for example, the El Nino-Southern Oscillation (ENSO) modulate the frequency and intensity of MHWs in SCS via teleconnection mechanisms (Zhang et al., 2025). For instance, during ENSO, PSL and wind fields anomalies can influence the intensity and position of the Western Pacific Subtropical High (WPSH), thereby altering the surface thermal structure of the SCS. This ultimately leads to localized SSTA, increasing the probability of MHW.

Although the RF model can quantify the global importance of predictors, it is difficult to provide local interpretability and the dependencies among variables. This study introduces the SHapley Additive exPlanations (SHAP) method, which enables both global and local model interpretation. Fig. 7 displays a SHAP summary plot visualizing how predictors contribute to model outputs, thereby illuminating the model's decision mechanisms (Li et al., 2022; Li et al., 2022).

For 7-day forecast horizons, the ranking of the top 15 most important predictors identified by SHAP is highly consistent with the feature importance rankings from the RF model. Although slight variations exist in the ranking of specific variables across other forecast horizons, the composition

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

of the top 15 key variables remains relatively stable, further validating their predictive importance. In contrast to global importance scores, SHAP provides more intuitive visualizations of variable impacts. Among key predictors except the two wind components (UAS and VAS), high values of most variables show a positive correlation with positive SHAP values—meaning their increase promotes MHW prediction. This implies the model is more likely to identify MHWs when these variables are elevated. For example, higher SSTA typically indicates warmer sea surface temperatures, which are closely associated with the MHW formation.

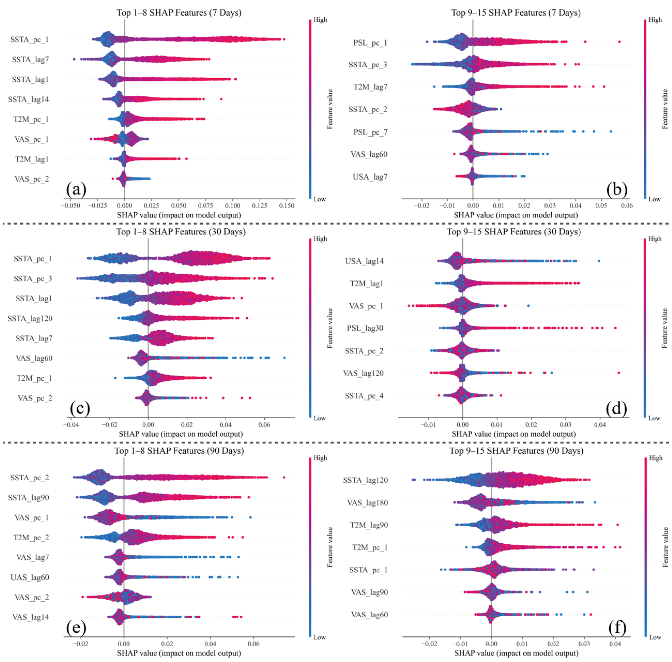

Fig.7. SHAP feature importance distribution of the RF model, showing the contribution degree and direction of the top 15 predictor variables across different forecast lead times (7-day, 30-day, and 90-day). (a) and (b) correspond to 7-day, (c) and (d) to 30-day, and (e) and (f) to 90-day. Each data point represents a sample, with the x-axis denoting the SHAP value (degree and direction of feature contribution to model predictions) and the y-axis listing predictor variable names. The color of each data point indicates the degree of the corresponding feature value, with red denoting higher values and blue indicating lower values. The horizontal position of data points represents the degree of influence of the feature on model predictions. Larger SHAP values on the right indicate a greater contribution of the feature to predicting MHWs, while smaller or negative SHAP values on the left signify stronger inhibitory effects.

Warming in the Indian Ocean may trigger an anti-Walker circulation between the tropical Indian Ocean and western Pacific, of which enhances the Philippine Anticyclone (PAC). Such large-scale circulation changes weaken convective activity in the northwestern Pacific, promoting westward extension of the WPSH toward China's southern coast (Liang et al., 2023). The westward expansion of WPSH typically suppresses easterly anomalies induced by the SCS anticyclonic circulation, leading to weakening of the southwest monsoon. This further diminishes upwelling and cold surge phenomena, creating favorable conditions for MHW formation (Xie et al., 2003; Song et al., 2023).

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

On the other hand, lower wind speeds reduce sea surface evaporation, decreasing heat release from the ocean to the atmosphere and exacerbating oceanic heat storage—processes that facilitate MHW development. Meanwhile, La Niña-type tropical Pacific patterns enhance westward ocean currents via anomalous easterlies, transporting warm water to the western Pacific and SCS. These results demonstrate that the SHAP method not only enhances model interpretability but also further validates that the model can identify key factors with physical significance.

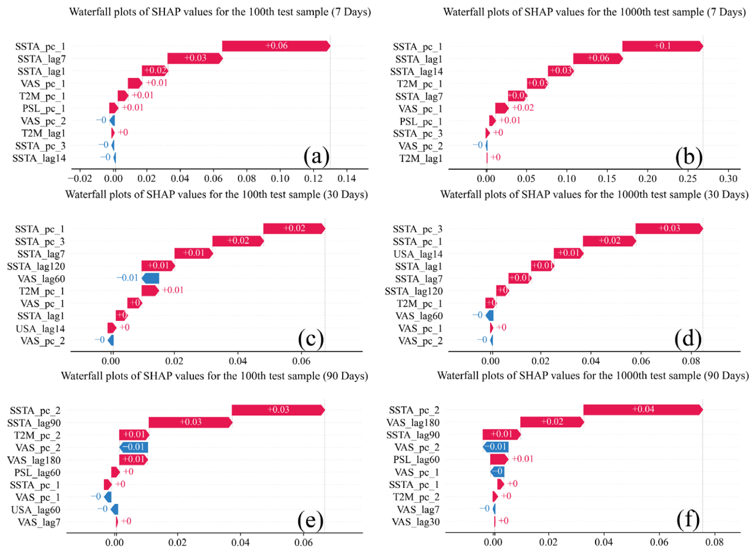

Fig.8. SHAP waterfall plots of the RF model, showing the contribution degree and direction of the top 10 important predictor variables for the 100th and 1000th test samples across different forecast lead times. (a), (c), and (e) represent the 100th sample for 7-day, 30-day, and 90-day, respectively, while (b), (d), and (f) correspond to the 1000th sample. Each bar denotes a predictor, with its length and arrow direction indicating the contribution deggre and sign to MHW occurrence probability: red bars extending to the right signify positive contributions (increasing MHW probability), blue bars extending to the left denote negative contributions (decreasing probability), and the numerical values beside bars represent the absolute contribution degree.

SHAP not only evaluates global feature importance, but also offers strong local interpretability, enabling in-depth revelation of the model's decision basis for individual prediction samples. To analyze how key predictive factors influence specific sample predictions, the 100th and 1000th test samples were selected as representative cases. Both samples belong to MHWs and were successfully predicted by the model in all forecast lead times.

Fig. 8 shows SHAP waterfall plots illustrating the local interpretability results of the top 10 features for these two samples in 7-day, 30-day, and 90-day forecast lead time. Left panels (a, c, e) correspond to the 100th sample, while right panels (b, d, f) correspond to the 1000th sample. Red arrows mean positive contributions, blue arrows denote negative impacts, and arrow length represents the degree of SHAP values (i.e., the contribution degree of features to prediction results). The plots show that SSTA and T2M consistently rank among the top features in all forecast lead

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

times, exerting significant positive impacts on model predictions. As shown in the figure, SSTA and T2M consistently rank among the top predictors in all forecast lead times, exerting significant positive impacts on model predictions. In particular, the principal components and lagged terms of SSTA exhibit the highest contribution values, indicating their dominant role in MHW prediction. This aligns with MHW formation mechanisms: higher SSTA and T2M signify warm anomalies in the SCS, providing thermal support for MHWs. Additionally, most SHAP values of wind field variables (UAS, VAS) are negative, suggesting that lower values promote MHW prediction. This mechanism is evident in multiple forecast lead times, with the negative contribution of wind field lagged terms being particularly prominent in 90-day predictions.

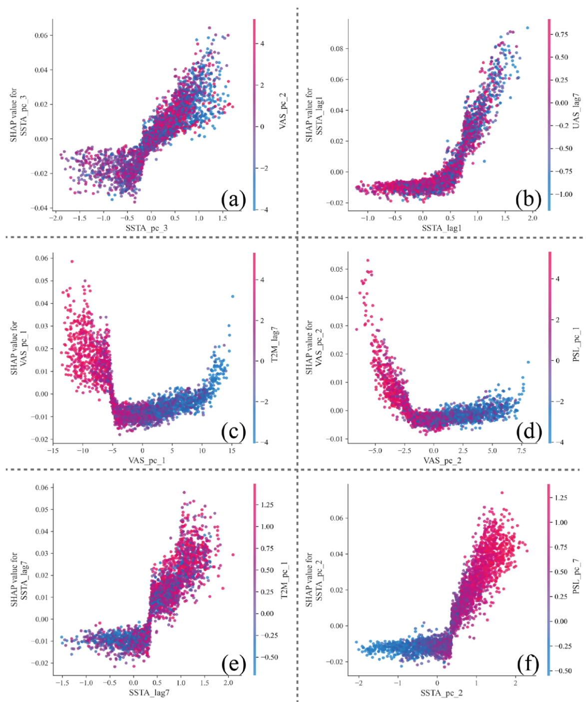

Fig.9. SHAP two-factor dependency plots for 7-day predictions of the RF model, demonstrating the interaction of six groups of predictor variables in South China Sea MHW prediction. In each subplot, the x-axis represents the actual value of the dominant variable in the group, and the y-axis represents the corresponding SHAP value (degree of contribution). The color of data points indicates the magnitude of the second variable: red for high values and blue for low values.

SHAP values can not only reveal the global and local contributions of individual predictors but also analyze the interaction effects between different predictors. To explore the dependency relationships and nonlinear response mechanisms among variables, this study focused on the 7-day forecast lead time and generated SHAP dependency plots, as illustrated in Fig. 9.

Fig. 9 illustrates the relationship between SHAP values and input variable values. The figure uses color to represent the values of another interacting variable, revealing the synergistic effects between

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

predictor variables. For example, in Fig. 9(a) and 9(b), the principal components and lagged terms of SSTA (SSTA_pc_3, SSTA_lag1) exhibit nonlinear response features. When SSTA values are low, their SHAP values are negative and inhibit MHW prediction. Once exceeding a threshold (approximately 0) SHAP values rise rapidly and turn positive. This not only indicates that a higher SSTA drives the occurrence of MHW, but also serves as an illustration of the pronounced sensitivity exhibited by the model towards variations in SSTA. Wind-related features show contrasting patterns. As shown in Figures 9(c) and 9(d), the principal components of wind fields (VAS_pc_1, VAS_pc_2) have negative SHAP values at high wind speeds, inhibiting model predictions. At low wind speeds, SHAP values tend to be positive, suggesting that lower wind speeds facilitate oceanic heat accumulation and MHW formation. Meanwhile, color gradients indicate that at the same SSTA or VAS values, the values of the other variable significantly affect SHAP values, reflecting strong interaction effects between SSTA and wind fields. The Fig. 9(e) and 9(f) further reveal the synergistic influence mechanisms of SSTA with T2M and PSL. In Fig. 9(e), the SHAP value of SSTA_lag7 rises nonlinearly with its own increase, and the value of T2M_pc_1 significantly modulates it, indicating that rising air temperature can enhance the influence of SSTA. The Fig. 9(f) shows that when PSL_pc_2 is high, the positive contribution of SSTA_pc_2 is amplified, which may reflect the enhanced role of WPSH westward extension in MHW formation.

In summary, SHAP dependency plots reveal nonlinear interaction relationships among key variables, where SSTA and wind field components form the core factor combination for MHW prediction. The study finds that SSTA has a stronger predictive power in short-term forecasts, with its importance gradually diminishing as the forecast lead times extends. In contrast, the importance of wind fields (especially VAS) grows in long-term forecasts, aligning with their physical mechanisms of regulating oceanic heat fluxes and influencing upwelling. These results not only enhance the interpretability of models but also provide a theoretical basis for understanding MHW formation mechanisms and prediction.

5 Conclusion and Discussion

This study tested a method for predicting MHWs in the SCS using non-homologous training and testing datasets. CMIP6 ensemble was used to train multiple machine learning models and combined observational data for MHW prediction. The results showed that short-term prediction accuracy is higher than that of long-term forecasts. In both lead times, the machine learning models demonstrated superior predictive skill compared with baseline methods, achieving performance comparable to or exceeding that of other climate simulation models.

Then, the relative importance of predictors was accessed by using feature importance analysis in the Random Forest model and SHAP values. These predictors includes both SST-related variables and atmospheric factors, so as to enhance new model's physical interpretability. The analysis revealed that SST and wind fields are the two most critical factors influencing MHWs. SST variables dominated in short-term predictions, while wind-related features became increasingly important in mid- to long-term forecasts. Compared with traditional models trained solely on observational data, models trained on extensive climate simulation outputs are better equipped to overcome sampling limitations and represent complex nonlinear dynamics. Integrating multiple climate models into machine learning training further enhances predictive performance and improves model generalizability.

Among the machine learning models tested in this study, Random Forest (RF) model shows the best prediction performance. Its advantage may come from its strong interpretability. It also uses fewer parameters and is less sensitive to parameter settings. This makes it suitable for handling the number of variables involved in this study. However, as more complex and diverse features may be introduced in the future, it is still necessary to explore deep learning models with stronger nonlinear fitting capabilities. Moreover, this study focuses only on the SCS. In the future, the method can be extended to other ocean regions or even the global scale. This will help to learn the characteristics and evolution patterns of MHWs in different regions. It will also provide more comprehensive support for the development of a global marine heatwave early warning system.

Acknowledgments

This study was supported in part by the Innovation Team Project of the General University in the Guangdong Province, China (Grant number 2024KCXTD042).

Author contributions

Peihao Yang: Writing-original draft, Software, Validation, Methodology.
Guodong Ye: Supervision, Conceptualization, Methodology, Investigation, Resources, Project administration, Writing-review & editing.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare that they have no competing financial interests or personal relationships that may have influenced the work reported in this study.

Data availability

These datasets are publicly available. Sea surface temperature data are taken from the NOAA dataset

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

(https://www.esrl.noaa.gov/psd). SST-related meteorological data taken from the ERA5 dataset (https://cds.climate.copernicus.eu).

References

Amaya, D. J., Alexander, M. A., Capotondi, A., Deser, C., Karnauskas, K. B., Miller, A. J. and Mantua, N. J.: Are long-term changes in mixed layer depth influencing North Pacific marine heatwaves?, Bulletin of the American Meteorological Society, 102, S59–S66, 2021.

Amaya, D. J., Miller, A. J., Xie, S. P. and Kosaka, Y.: Physical drivers of the summer 2019 North Pacific marine heatwave, Nature Communications, 11, 1903, 2020.

Barnston, A. G. and Smith, T. M.: Specification and prediction of global surface temperature and precipitation from global SST using CCA, Journal of Climate, 9, 2660–2697, 1996.

Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M. and Vialard, J.: ENSO representation in climate models, from CMIP3 to CMIP5, Climate Dynamics, 42, 1999–2018, 2014.

Cabaneros, S. M., Calautit, J. K. and Hughes, B. R.: A review of artificial neural network models for ambient air pollution prediction, Environmental Modelling & Software, 119, 285–304, 2019.

Chen, K., Gawarkiewicz, G., Kwon, Y. O. and Zhang, W. G.: The role of atmospheric forcing versus ocean advection during the extreme warming of the Northeast US continental shelf in 2012, Journal of Geophysical Research: Oceans, 120, 4324–4339, 2015.

Chen, Z., Ding, S. X., Peng, T., Yang, C. and Gui, W.: Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms, IEEE Transactions on Industrial Electronics, 65, 1559–1567, 2017.

Di Lorenzo, E. and Mantua, N.: Multi-year persistence of the 2014/15 North Pacific marine heatwave, Nature Climate Change, 6, 1042–1047, 2016.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J. and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development, 9, 1937–1958, 2016.

Feng, Y., Bethel, B. J., Dong, C., Zhao, H., Yao, Y. and Yu, Y.: Marine heatwave events near Weizhou Island, Beibu Gulf in 2020 and their possible relations to coral bleaching, Science of the Total Environment, 823, 153414, 2022.

Frölicher, T. L. and Laufkötter, C.: Emerging risks from marine heat waves, Nature Communications, 9, 650, 2018.

Gan, J., Li, H., Curchitser, E. N. and Haidvogel, D. B.: Modeling South China Sea circulation, response to seasonal forcing regimes, Journal of Geophysical Research: Oceans, 111, C06034, 2006.

Gao, G., Yao, B., Li, Z., Duan, D. and Zhang, X.: Forecasting of sea surface temperature in eastern tropical Pacific by a hybrid multiscale spatial–temporal model combining error correction map, IEEE Transactions on Geoscience and Remote Sensing, 62, 1–22, 2024.

Ham, Y.-G., Kim, J.-H. and Luo, J.-J.: Deep learning for multi-year ENSO forecasts, Nature, 573, 568–572, 2019.

Ham, Y.-G., Kim, J.-H., Min, S.-K., Kim, D., Li, T., Timmermann, A. and Stuecker, M. F.: Anthropogenic fingerprints in daily precipitation revealed by deep learning, Nature, 622, 301–307, 2023.

Han, W., Zhang, L., Meehl, G. A., Kido, S., Tozuka, T., Li, Y., McPhaden, M. J., Hu, A., Cazenave, A., Rosenbloom, N., Strand, G., West, B. J. and Xing, W.: Sea level extremes and compounding marine heatwaves in coastal Indonesia, Nature Communications, 13, 6410, 2022.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S. and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, 2020.

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C. J., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Feng, M., Holbrook, N. J., Moore, P. J., Scannell, H. A., Sen Gupta, A. and Wernberg, T.: A hierarchical approach to defining marine heatwaves, Progress in Oceanography, 141, 227–238, 2016.

Hobday, A. J., Oliver, E. C. J., Sen Gupta, A., Benthuysen, J. A., Burrows, M. T., Donat, M. G., Holbrook, N. J., Moore, P. J., Thomsen, M. S., Wernberg, T. and Smale, D. A.: Categorizing and naming marine heatwaves, Oceanography, 31, 162–173, 2018.

Holbrook, N. J., Sen Gupta, A., Oliver, E. C. J., Hobday, A. J., Benthuysen, J. A., Scannell, H. A., Smale, D. A. and Wernberg, T.: Keeping pace with marine heatwaves, Nature Reviews Earth & Environment, 1, 482–493, 2020.

Hu, J. and Szymczak, S.: A review on longitudinal data analysis with random forest, Briefings in Bioinformatics, 24, bbad002, 2023.

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H, Menne, M. J., Smith, T. M., Vose, R. S. and Zhang, H.-M.: Extended reconstructed sea surface temperature, version 5 (ERSSTv5), upgrades, validations, and intercomparisons, Journal of Climate, 30, 8179–8205, 2017.

Jeffree, J., Hogg, A. M. C. C., Morrison, A. K., Solodoch, A., Stewart, A. L. and McGirr, R.: GRACE satellite observations of Antarctic Bottom Water transport variability, Journal of Geophysical Research: Oceans, 129, e2024JC020990, 2024.

Kim, J., Oh, H. S., Lim, Y. and Kang, H. S.: Seasonal precipitation prediction via data-adaptive principal component regression, International Journal of Climatology, 37, 75–86, 2017.

Lee, T., Hobbs, W. R., Willis, J. K., Halkides, D., Fukumori, I., Armstrong, E. M., Hayashi, A. K., Liu, W. T., Patzert, W. and Wang, O.: Record warming in the South Pacific and western Antarctica associated with the strong central-Pacific El Niño in 2009–10, Geophysical Research Letters, 37, L19704, 2010.

Li, L., Qiao, J., Yu, G., Wang, L., Li, H. Y., Liao, C. and Zhu, Z.: Interpretable tree-based ensemble

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

model for predicting beach water quality, Water Research, 211, 118078, 2022.

Li, Z., Liu, F., Yang, W., Peng, S. and Zhou, J.: A survey of convolutional neural networks, analysis, applications, and prospects, IEEE Transactions on Neural Networks and Learning Systems, 33, 6999–7019, 2021.

Li, Z.: Extracting spatial effects from machine learning model using local interpretation method, an example of SHAP and XGBoost, Computers, Environment and Urban Systems, 96, 101845, 2022.

Liang, C. M., Liang, Z., Zhou, S. W., Shen, X. Y., Huang, C. R., Ding, Y. H., Liu, Y. J., Yao, H. X. and Zhou, K. X.: Rapid increase in warm–wet compound extreme events with high health risks in southern China, joint influence of ENSO and the Indian Ocean, Advances in Climate Change Research, 14, 856–865, 2023.

Lin, A., Wu, H., Liang, G., Cardenas-Tristan, A., Wu, X., Zhao, C. and Li, D.: A big data-driven dynamic estimation model of relief supplies demand in urban flood disaster, International Journal of Disaster Risk Reduction, 49, 101682, 2020.

Liu, K., Xu, K., Zhu, C. and Liu, B.: Diversity of marine heatwaves in the South China Sea regulated by ENSO phase, Journal of Climate, 35, 877–893, 2022.

Martinez-Sosa, P., Tierney, J. E., Stefanescu, I. C., Crampton-Flood, E. D., Shuman, B. N. and Routson, C.: A global Bayesian temperature calibration for lacustrine brGDGTs, Geochimica et Cosmochimica Acta, 305, 87–105, 2021.

Mo, S., Chen, T., Chen, Z., Zhang, W. and Li, S.: Marine heatwaves impair the thermal refugia potential of marginal reefs in the northern South China Sea, Science of the Total Environment, 825, 154100, 2022.

Niazkar, M., Menapace, A., Brentan, B., Piraei, R., Jimenez, D., Dhawan, P. and Righetti, M.: Applications of XGBoost in water resources engineering, a systematic literature review (Dec 2018–May 2023), Environmental Modelling & Software, 174, 105971, 2024.

Oliver, E. C. J., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., Benthuysen, J. A., Feng, M., Sen Gupta, A., Hobday, A. J., Holbrook, N. J., Perkins-Kirkpatrick, S. E., Scannell, H. A., Straub, S. C. and Wernberg, T.: Longer and more frequent marine heatwaves over the past century, Nature Communications, 9, 1324, 2018.

Park, S., Im, J., Jang, E. and Rhee, J.: Drought assessment and monitoring through blending of multi-sensor indices using machine learning approaches for different climate regions, Agricultural and Forest Meteorology, 216, 157–169, 2016.

Perkins-Kirkpatrick, S. E., King, A. D., Cougnon, E., Grose, M. R., Oliver, E. C. J., Holbrook, N., Lewis, S. C. and Pourasghar, F.: The role of natural variability and anthropogenic climate change in the 2017/18 Tasman Sea marine heatwave, Bulletin of the American Meteorological Society, 100, S105–S110, 2019.

Qi, R., Zhang, Y., Du, Y. and Feng, M.: Characteristics and drivers of marine heatwaves in the western equatorial Indian Ocean, Journal of Geophysical Research: Oceans, 127, e2022JC018732, 2022.

Rana, S., Renwick, J., McGregor, J. and Singh, A.: Seasonal prediction of winter precipitation

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

anomalies over Central Southwest Asia, a canonical correlation analysis approach, Journal of Climate, 31, 727–741, 2018.

Roy, A. and Chakraborty, S.: Support vector machine for structural reliability analysis, a review, Reliability Engineering & System Safety, 233, 109126, 2023.

Saranya, J. S., Roxy, M. K., Dasgupta, P. and Anand, A.: Genesis and trends in marine heatwaves over the tropical Indian Ocean and their interaction with the Indian summer monsoon, Journal of Geophysical Research: Oceans, 127, e2021JC017427, 2022.

Sattari, M. T., Feizi, H., Samadianfard, S., Falsafian, K. and Salwana, E.: Estimation of monthly and seasonal precipitation, a comparative study using data-driven methods versus hybrid approach, Measurement, 173, 108512, 2021.

Séférian, R., Berthet, S., Yool, A., Palmiéri, J., Bopp, L., Tagliabue, A., Kwiatkowski, L., Aumont, O., Christian, J., Dunne, J., Gehlen, M., Ilyina, T., Li, H., Long, M. C., Luo, J. Y., Nakano, H., Romanou, A., Schwinger, J., Stock, C., Santana-Falcón, Y., Takano, Y., Tjiputra, J., Tsujino, H., Watanabe, M., Wu, T., Wu, F. and Yamamoto, A.: Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6, Current Climate Change Reports, 6, 95–119, 2020.

Sen Gupta, A., Thomsen, M., Benthuysen, J. A., Hobday, A. J., Oliver, E. C. J., Alexander, L. V., Burrows, M. T., Donat, M. G., Feng, M., Holbrook, N. J., Perkins-Kirkpatrick, S., Moore, P. J., Rodrigues, R. R., Scannell, H. A., Taschetto, A. S., Ummenhofer, C. C., Wernberg, T. and Smale, D. A.: Drivers and impacts of the most extreme marine heatwave events, Scientific Reports, 10, 19359, 2020.

Slater, L. J., Arnal, L., Boucher, M.-A., Chang, A. Y. Y., Moulds, S., Murphy, C., Nearing, G., Shalev, G., Shen, C., Speight, L., Villarini, G., Wilby, R. L., Wood, A. and Zappa, M.: Hybrid forecasting, blending climate predictions with AI models, Hydrology and Earth System Sciences, 27, 1865–1889, 2023.

Smale, D. A., Wernberg, T., Oliver, E. C. J., Thomsen, M., Harvey, B. P., Straub, S. C., Burrows, M. T., Alexander, L., Benthuysen, J. A., Donat, M. G., Feng, M., Hobday, A. J., Holbrook, N. J., Perkins-Kirkpatrick, S. E., Scannell, H. A., Sen Gupta, A., Payne, B. L. and Moore, P. J.: Marine heatwaves threaten global biodiversity and the provision of ecosystem services, Nature Climate Change, 9, 306–312, 2019.

Song, Q., Yao, Y. and Wang, C.: Response of future summer marine heatwaves in the South China Sea to enhanced western Pacific subtropical high, Geophysical Research Letters, 50, e2023GL103667, 2023.

Tan, H. J., Cai, R. S. and Wu, R. G.: Summer marine heatwaves in the South China Sea, trend, variability and possible causes, Advances in Climate Change Research, 13, 323–332, 2022.

Tang, S., Chen, R. and Chen, Z.: Spatial diversity of the summer marine heat waves in the South China Sea, Journal of Climate, 38, 1051–1065, 2025.

Tao, L. J., Zhang, R. H. and Gao, C.: Initial error-induced optimal perturbations in ENSO predictions, as derived from an intermediate coupled model, Advances in Atmospheric Sciences, 34, 791–803, 2017.

Van der Ploeg, T., Austin, P. C. and Steyerberg, E. W.: Modern modelling techniques are data hungry,

https://doi.org/10.5194/egusphere-2026-262
Preprint. Discussion started: 30 January 2026

a simulation study for predicting dichotomous endpoints, BMC Medical Research Methodology, 14, 1–13, 2014.

Waliser, D. E., Lau, K. M., Stern, W. and Jones, C.: Potential predictability of the Madden–Julian oscillation, Bulletin of the American Meteorological Society, 84, 33–50, 2003.

Wang, H., Liang, Q., Hancock, J. T. and Khoshgoftaar, T. M.: Feature selection strategies, a comparative analysis of SHAP-value and importance-based methods, Journal of Big Data, 11, 44, 2024.

Xiao, F., Wang, D., Zeng, L., Liu, Q. and Zhou, W.: Contrasting changes in the sea surface temperature and upper ocean heat content in the South China Sea during recent decades, Climate Dynamics, 53, 1597–1612, 2019.

Xie, S. P., Xie, Q., Wang, D. and Liu, W. T.: Summer upwelling in the South China Sea and its role in regional climate variations, Journal of Geophysical Research: Oceans, 108, 3261, 2003.

Yang, P. and Ye, G.: Tropical cyclone track prediction model for multidimensional features and time differences series observation, Alexandria Engineering Journal, 111, 432–445, 2025.

Yao, Y. and Wang, C.: Variations in summer marine heatwaves in the South China Sea, Journal of Geophysical Research: Oceans, 126, e2021JC017792, 2021.

Yao, Y., Wang, J., Yin, J. and Zou, X.: Marine heatwaves in China's marginal seas and adjacent offshore waters, past, present, and future, Journal of Geophysical Research: Oceans, 125, e2019JC015801, 2020.

Yu, Y., Si, X., Hu, C. and Zhang, J.: A review of recurrent neural networks, LSTM cells and network architectures, Neural Computation, 31, 1235–1270, 2019.

Zhang, L., Delworth, T. L., Cooke, W. and Yang, X.: Natural variability of Southern Ocean convection as a driver of observed climate trends, Nature Climate Change, 9, 59–65, 2019.

Zhang, N., Lan, J. and Dong, C.: Subsurface heatwaves and cold spells in the South China Sea regulated by ENSO, role of the South China Sea throughflow, Geophysical Research Letters, 52, e2025GL114692, 2025.

Zhang, X., Hao, Z., Singh, V. P., Zhang, Y, Feng, S., Xu, Y. and Hao, F.: Drought propagation under global warming, characteristics, approaches, processes, and controlling factors, Science of the Total Environment, 838, 156021, 2022.

Zhao, Y., Chen, M., Chung, T. H., Chan, L. L. and Qiu, J. W.: The 2022 summer marine heatwaves and coral bleaching in China's Greater Bay Area, Marine Environmental Research, 189, 106044, 2023.

Zhu, L. and Aguilera, P.: Evaluating variations in tropical cyclone precipitation in Eastern Mexico using machine learning techniques, Journal of Geophysical Research: Atmospheres, 126, e2021JD034604, 2021.