

Article

An Ensemble-Based Machine Learning Model for Estimation of Subsurface Thermal Structure in the South China Sea

Jifeng Qi ^{1,2,3}, Chuanyu Liu ^{1,2,3}, Jianwei Chi ^{4,5}, Delei Li ^{1,2,3}, Le Gao ^{1,2,3}  and Baoshu Yin ^{1,2,3,6,*}

¹ CAS Key Laboratory of Ocean Circulation and Waves, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China; jfqi@qdio.ac.cn (J.Q.); chuanyu.liu@qdio.ac.cn (C.L.); delei@qdio.ac.cn (D.L.); gaoale@qdio.ac.cn (L.G.)

² Pilot National Laboratory for Marine Science and Technology (Qingdao), Qingdao 266237, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

⁴ State Key Laboratory of Tropical Oceanography, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou 510301, China; chijianwei@scso.ac.cn

⁵ Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China

⁶ CAS Engineering Laboratory for Marine Ranching, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China

* Correspondence: bsyin@qdio.ac.cn

Abstract: Reconstructing the vertical structures of the ocean from sea surface information is of great importance for ocean and climate studies. In this study, an ensemble machine learning (Ens-ML) model is proposed to retrieve ocean subsurface thermal structure (OSTS) by using satellite-derived sea surface data and Argo data in the South China Sea (SCS). The input data include sea surface height (SSH), sea surface temperature (SST), sea surface salinity (SSS), sea surface wind (SSW), and geographic information (including longitude and latitude). We select three stable machine learning models, namely, extreme gradient boosting (XGBoost), RandomForest and light gradient boosting machine (LightGBM) as our benchmark models, and then use an artificial neural network (ANN) technique to combine outputs from the three individual models. The proposed Ens-ML model using sea surface data only by SSH, SST, SSS, and SSW performs less satisfactorily than that considering the contribution of geographical information, indicating that the geographical information is essential to estimate the OSTs accurately. The estimated OSTs from the Ens-ML model are compared with Argo data. The results show that the proposed Ens-ML model can accurately estimate the OSTs (upper 1000 m) in the SCS, which is relatively more accurate and precise than the individual models. The performance of the Ens-ML model also varies with season, and better estimation is obtained in winter, which is probably due to stronger mixing and weaker stratification. This study shows the great potential and advantage of the multi-model ensemble of machine learning algorithm for the ocean's interior information retrieving, showing great potential in expanding the scope of ocean observations.



Citation: Qi, J.; Liu, C.; Chi, J.; Li, D.; Gao, L.; Yin, B. An Ensemble-Based Machine Learning Model for Estimation of Subsurface Thermal Structure in the South China Sea. *Remote Sens.* **2022**, *14*, 3207.

<https://doi.org/10.3390/rs14133207>

Academic Editors:

Ferdinando Nunziata, Gang Zheng,
Weizeng Shao and Juhong Zou

Received: 19 May 2022

Accepted: 1 July 2022

Published: 4 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Ocean temperature, one of the most important variables of seawater, plays a significant role in marine disasters, marine ecosystems, ocean dynamics, and climate changes [1,2], for example, it is closely related to the formation of many high impact thermos-dynamical ocean-atmospheric processes, such as the marine heatwaves [3,4], thermocline layer formation [5], El Niño evolution [6,7], or deep-water formation [8,9]. To better understand the role of ocean temperature in marine disaster and ecosystems, ocean circulation and climate changes, it is essential to identify the ocean subsurface thermal structure (OSTS) accurately.

As the marginal sea of the western Pacific, the South China Sea (SCS) lies under the influence of the Western Pacific Warm Pool [10,11]. It has an extended continental shelf

in the south and extends deeper than 5000 m in the central region (Figure 1a). Extended continental shelves exist along the north boundary and southwest boundary, while the deeper water is oriented in a southwest–northeast direction across the central part of the SCS [12]. Due to its unique geographical location and complicated dynamical features, the temperature structures in the SCS have remarkable characteristics associated with El Niño–Southern Oscillation (ENSO) [13], Asian monsoon and Pacific western boundary current system [14,15], which have a far-reaching influence on the regional and global climate [16,17]. In the climatological mean, sea surface temperature (SST) in the SCS shows a chiefly southwest–northeast oriented pattern, which is gradually increasing southward (Figure 1b). Previous studies confirmed that ocean temperature plays a vital role in the ecosystems, dynamics and thermodynamics of the SCS. For example, ocean temperature variability can significantly influence the occurrence of marine heatwaves in the SCS [18]. In addition, ocean temperature can also play a role in the processes of air–sea interactions in the SCS, which can exert noticeable impacts on the East Asian climate [19]. Despite the importance of the ocean temperature structures in the SCS, it remains a challenging problem for oceanographers and climatologists to accurately estimate them. Current observations of OSTs in the SCS are sparse and insufficient, which significantly constrain the studies of ocean processes. Therefore, the most commonly used methods to estimate the OSTs are numerical simulation, numerical-based data assimilation and dynamical modeling [20,21]; however, these dynamical models, due to simulating full aspects of the ocean thermal and dynamical processes that are governed by a suit of equations, are very time and computationally consuming. How to accurately estimate the OSTs with an acceptable computational resource is an active area of research.

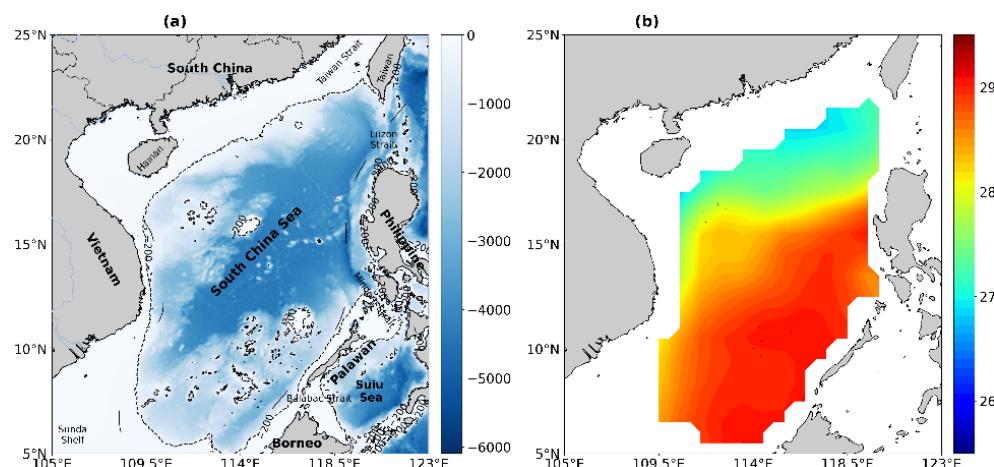


Figure 1. (a) Bathymetry (m), and (b) long-term mean (from January 2010 to December 2019) SST from Argo data in the South China Sea.

In recent decades, the latest technological advances in remote sensing have significantly improved our understanding of ocean circulation and dynamics by providing well-sampled surface observations, such as sea surface height (SSH), SST, sea surface salinity (SSS), and sea surface wind (SSW). Several early studies suggested that many ocean subsurface phenomena have surface manifestations, therefore, satellite-derived sea surface data can be used to reconstruct the ocean interior [22–26]. In this regard, extensive efforts have been made to reconstruct ocean interior structures via surface information from satellite data [27–31]. Methods for retrieving ocean interior structures from sea surface data are often based on linear regression of single or multiple variables [32,33], statistical and dynamic methods [34], or machine learning methods [24,35,36]. Khedouri et al. [37] used purely statistical relationships to estimate OSTs from sea surface data, such as sea surface dynamic height and SST. Based on the empirical orthogonal function (EOF), DeWitt [38] presented a statistical method to develop the relationship between dynamic height and am-

plitudes of the first two vertical modes. They pointed out that sea surface dynamic height is a key parameter to estimate ocean temperature structure. Later, Carnes et al. [23] retrieved the ocean vertical temperature and salinity structures based on a regression relationship between sea surface height and ocean vertical temperature structures. Chu et al. [28] developed a parametric model to retrieve OSTs from SST observations. Based on a “gravest empirical mode” (GEM), Watts et al. [39] found that the temperature and salinity structures between 150 and 3000 dbars in the south of Australia can be accurately reconstructed through GEM. Based on satellite altimetry data, Meijers et al. [34] used the GEM method to estimate ocean subsurface temperature and salinity structures in the Southern Ocean. Guinehut et al. [33] achieved good performance in the estimation of temperature and salinity structures based on satellite-derived data and in-situ measurements through linear regression methods. Su et al. [40] developed a geographically weighted regression model to retrieve the ocean subsurface temperature anomaly in the Indian Ocean, which outperforms the ordinary linear regression model. Yu et al. [41] proposed a ridge regression model to derive ocean thermal structure in the Bay of Bengal using satellite altimetry data and Argo in-situ data.

Recently, machine learning technology has been widely used in many fields to solve specific problems throughout academia, including fields of atmospheric and oceanic sciences [42]. Several early studies have used machine learning models to retrieve ocean interior structure from multisource surface data. For example, Ali et al. [24] adopted an artificial neural network (ANN) method to reconstruct the vertical temperature structure from sea surface data such as SST, SSS, wind stress, net radiation, and net heat flux. They suggested that the ANN approach has a good performance in estimating OSTs from sea surface data. Later, Wu et al. [35] used a self-organizing map (SOM) neural network to retrieve ocean subsurface temperature structures from sea surface data. Combined with EOF analysis, Chen et al. [43] used a SOM approach to retrieve the subsurface thermal structure via sea surface data in the Northwestern Pacific Ocean. Su et al. [44] have adopted a support vector machine (SVM) approach to retrieve the ocean vertical temperature structure anomaly in the Indian Ocean via sea surface data. This method has also been extended to estimate subsurface temperature anomalies on a global scale [45]. More recently, they tried other machine learning approaches such as RandomForest, extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM) to estimate the OSTs from sea surface data, and all of them have proven to be useful methods [26,46,47]. Han et al. [48] adopted a convolutional neural network (CNN) to retrieve OSTs in the Pacific using sea surface data, which yields a good performance. Lu et al. [49] adopted a method that combines a preclustering process and neural network approach to estimate subsurface thermal structures. They reported that the pre-clustered neural network method performs better than the same method without clustering. By using a stacked Long Short-Term Memory neural network (LSTM) method, a new model is proposed by Nardelli et al. [50] to estimate ocean hydrographic profiles in the North Atlantic Ocean. Recently, Meng et al. [51] adopted a CNN network to estimate subsurface and deep ocean temperature fields from satellite-derived sea surface data. They also proposed a scheme to improve the horizontal resolution of estimated temperature fields.

As suggested above, machine learning is a useful technology to estimate the subsurface thermal structure from sea surface data. The practical utility of estimates of OSTs derived from sea surface data will depend on the particular problem at hand. The performance of machine learning models mainly relies on algorithms and combinations of input parameters [31]. The optimal parameter combination used in machine learning models is found to depend on the kind of problem you wish to solve, the number of variables, the kind of model that would suit it best and so on. Most existing studies related to estimation of OSTs from sea surface data have focused on large-scale areas, such as global, Pacific Ocean, and Indian Ocean, but no related studies have been carried out in the SCS. Another thing to note here is that most of the previous studies were carried out using only individual machine learning models to estimate the OSTs, such as ANN, SOM, SVM, CNN,

XGBoost, RandomForest, and LightGBM, which may not make the perfect estimation for a given dataset due to their limitations. In recent years, the multi-model ensemble approach has been used to improve the overall accuracy of the prediction process, which usually yields better performance than individual models [52–54]. At present, the utilizations of the ensemble machine learning approach in the estimation of OSTs is still relatively rare. To resolve limitations in a single method, in this study, we used an ensemble machine learning (Ens-ML) model to estimate OSTs in the SCS. Based on previous studies, we chose three stable machine learning models (XGBoost, RandomForest, and LightGBM) as our basic machine learning model, and then we used an ANN to combine the outputs from each model. The remainder of the paper is organized as follows: Data and methods are described in Section 2, results are present in Section 3, and summary and discussion follow in Section 4.

2. Data and Methods

2.1. Data

The sea surface data used in this study are all from satellite observations, including SST, SSS, SSH and SSW (decomposed into eastward wind speed (USSW) and northward wind speed (VSSW) components). In addition, the geographic information including longitude (LON) and latitude (LAT) are also used in this study.

The monthly SST data are from the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation Sea Surface Temperature (OISST) version 2 products, which combine AVHRR, AMSR and in-situ observations [55]. The SSS data are obtained from Soil Moisture and Ocean Salinity (SMOS) Level-3 SSS product [56]. The SSH data measured by altimeters are obtained from Archiving, Validation and Interpretation of Satellite Oceanographic (AVISO) data center of CNES (Center National d'Etudes Spatiales) [57]. The SSW data are from the Cross-Calibrated Multi-Platform (CCMP) wind velocity product, which apply a variational analysis method to combine data from different sources to synthesize high-resolution surface wind data [58]. The Argo dataset used in this study is the monthly gridded Argo dataset produced by Roemmich and Gilson [59]. It has a horizontal resolution of $1^\circ \times 1^\circ$ and is interpolated to 58 depth levels (upper 2000 m) from January 2004 to present. In this study, the Argo data is used as labeled training data, as well as validating the accuracy and reliability of the results from models. Table 1 summarizes all the data used in this study. The Study area is located between 5~23°N latitude and 105~122°E longitude. It should be noted that all the above data have been processed into monthly averages and interpolated onto a grid with a resolution of 0.5° longitude and latitude grid with the same temporal and spatial coverage of the SCS. In addition, only the 2010–2019 time series for all variables have been considered.

Table 1. Summary of the data used in this study.

Index	Contents		
Study Area	South China Sea	105–122°E, 5–23°N	
Data	SST	2010–2019	NOAA (OISST)
	SSS	2010–2019	SMOS
	SSH	2010–2019	AVISO
	SSW (USSW, VSSW)	2010–2019	CCMP
	OSTS	2010–2019	RG-Argo
3D temperature field	Temporal and spatial resolution	monthly	$0.5^\circ \times 0.5^\circ$
	Vertical layers	2.5–1000 m	44 layers

2.2. Methods

Recently, the multi-model ensemble method has been widely studied, which has been demonstrated to yield more favorable results than individual machine learning models in many applications. In this study, we explore the use of a multi-model ensemble of machine learning approaches to estimate OSTs in the SCS. Among various ensemble learning methods, ANN is one of the most popular techniques to combine multiple models due to its effectiveness and simplicity. Generally, ensemble approaches are algorithm-independent, and impose no special restrictions on selection of the benchmark models. Based on the results of previous studies, we selected three stable machine learning models, XGBoost, RandomForest, and LightGBM as our basic model, all of which have been shown to have ability to estimate the OSTs [26,46,47,60].

In general, successful utilization of a machine learning application requires data under the categories: training data, validation data, and testing data. The training data are used to train the model. The validation data are used to validate the performance of the model during training and control overfitting almost simultaneously. In this study, all the surface input data from January 2010 to December 2018 were randomly split into two separate categories: 70% of the dataset for training and the remaining 30% for validation. As input, the sea surface data from January 2019 to December 2019 were used to predict the OSTs in the SCS. Out of the total 120 months of data, we used 75 months of data for training, 33 months of data for validation and 12 months of data for testing.

The flowchart of the proposed Ens-ML model for OSTs estimation in the SCS is shown in Figure 2. The Ens-ML model procedure involves two main steps: (1) training and validating the model, (2) estimating with model. The model was trained and validated using satellite-derived surface data and Argo labeled data before it was applied to reconstruct the OSTs. Based on previous studies and analyses, we selected SST, SSS, SSH, USSW, VSSW, and the geographical information, i.e., LON, and LAT, as independent input variables for machine learning models to estimate the OSTs in the SCS [24,26,47,51]. We first trained the three individual machine learning models (XGBoost, RandomForest, and LightGBM) to estimate the OSTs from sea surface data. After that, the outputs of the three machine learning models were combined using the ANN.

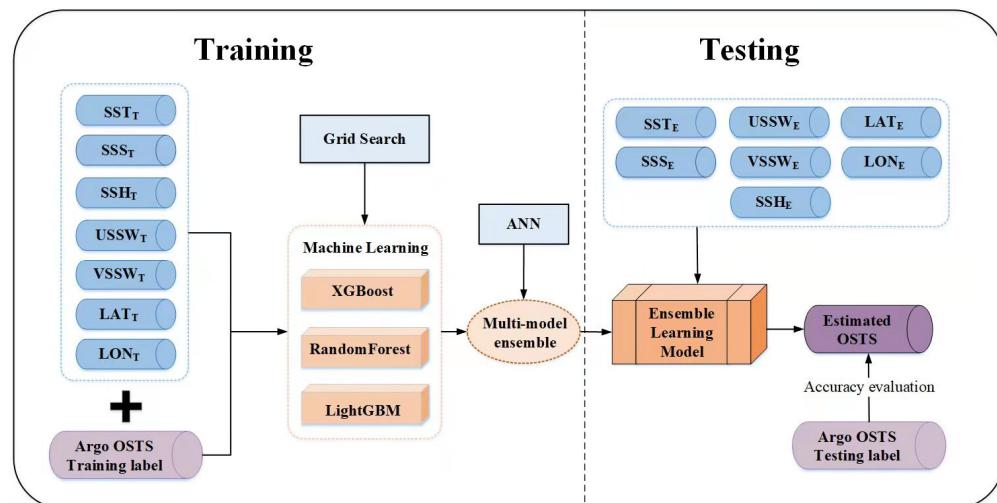


Figure 2. Flowchart for the ensemble machine learning (Ens-ML) model for ocean subsurface thermal structure estimation in the SCS.

Model parameter optimization, an important part of machine learning, has a significant influence on the performance of the models. Some of the most important parameters have to be carefully optimized to achieve good predictive performance. In this study, we applied the grid search method to obtain an optimal combination of model parameters for the machine learning models. The optimal parameter combinations are shown in Table 2.

Table 2. Optimal combination of parameters of models.

Model	Parameters
XGBoost	learning_rate = 0.3, n_estimators = 60, max_depth = 6, min_child_weight = 1, colsample_bytree = 1, colsample_bylevel = 1, subsample = 0.8, reg_lambda = 100
RandomForest	n_estimators = 150, max_depth = 21, min_samples_split = 70, min_samples_leaf = 3, max_features = 5, random_state = 10
LightGBM	num_leaves = 55, learning_rate = 0.01, n_estimators = 1000, max_depth = 8, min_child_samples = 20, feature_fraction = 0.8
ANN	number of neural network layers = 4, Residual layers = 2, learning rate = 0.002, batch_size = 1024

For model evaluation, the performance of the Ens-ML model was evaluated quantitatively using statistical measures, i.e., the root mean square error (RMSE hereafter) and coefficient of determination (R^2 hereafter). RMSE describes the specific values of the errors between the estimated values and observed values, while R^2 is used to assess how strong the linear relationship is between two variables. In the following section, we present a performance study on the Ens-ML model in the estimation of the OSTs in the SCS.

3. Results

With the optimal combination of parameters, we used three individual machine learning models (XGBoost, RandomForest, and LightGBM) and the Ens-ML model to separately estimate the OSTs in the SCS. First, we evaluated the performance and stability of the Ens-ML model by comparing with three individual models in terms of RMSE and R^2 , which were calculated at each depth levels based on the testing data set for 2019 over the SCS. The vertical distributions of RMSE and R^2 for the Ens-ML model and three individual models are shown in Figure 3.

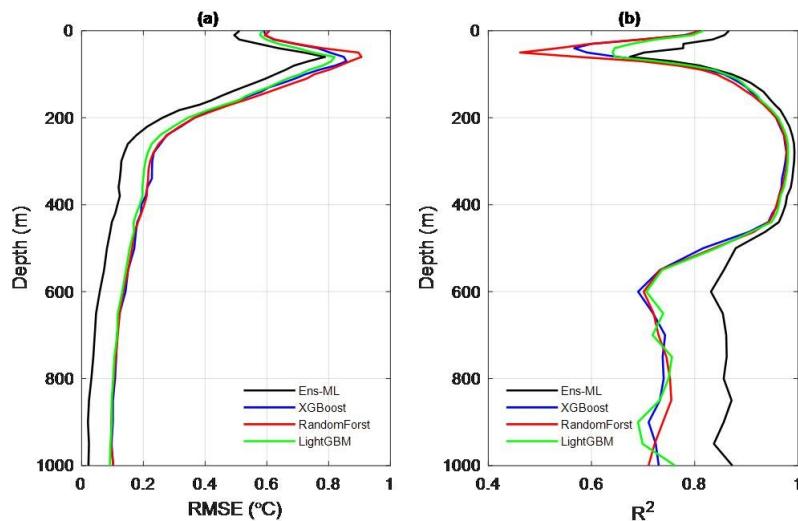


Figure 3. Comparison of the Ens-ML (black line) model with three individual machine learning models (XGBoost: blue line, RandomForest: red line, LightGBM: green line) for OSTs estimation at different depths based on (a) RMSE (°C) and (b) R^2 in 2019.

The overall trends for RMSE and R^2 of the Ens-ML model and three individual machine learning models are as follows: the RMSE increases from the surface to maximum values near 70 m depth, and then decreases with depth; whereas an opposite trend is seen in R^2 , which decreases from the surface to minimum values near 70 m, and then increases from 70 to 300 m, finally decreases from 300 to 1000 m. Our results show that, at about 70 m,

the RMSE is relative higher than other depths, whereas R^2 is relative lower than other depths. The depth of 70 m is approximately the depth of the thermocline layer, where the temperature decreases rapidly with depth, which leads to the difficulty of estimation of the OSTs [61]. In contrast, the RMSE values of the Ens-ML model are obviously smaller than those of the three individual models at different depths, while the R^2 values of the Ens-ML model are larger than those of the three individual models (Figure 3). Obviously, the Ens-ML model yields the highest overall accuracy indicated by the lowest RMSE ($0.31\text{ }^\circ\text{C}$, averaged over all depths) and highest R^2 (0.89) (Table 3). The second and third are the LightGBM model and the XGBoost model, respectively. The RandomForest model yields the lowest estimation accuracy. It should be pointed out that the gaps of R^2 below 450 m between the Ens-ML model and three individual models are larger than the upper layer. This indicates that the Ens-ML model, when compared with the three individual models, tends to perform better in the lower layers than in the upper layers.

Table 3. The performance comparison of different models for OSTs estimation per calculated yearly mean RMSE ($^\circ\text{C}$) and R^2 Values.

OSTS Estimation Models	RMSE	R^2
Ens-ML	0.31	0.89
XGBoost	0.39	0.83
RandomForest	0.40	0.83
LightGBM	0.37	0.84

Another important issue in the estimation is the selection of input variables for models. Early study has suggested that geographic information can help to improve estimation accuracy of the OSTs [47]. To examine the influence of geographic information on the estimation of the OSTs in the SCS, we set up two groups of experiments (cases 1 and 2) with different input parameter combinations to examine whether geographic information has an influence on the performance of the Ens-ML model (Table 4). In the first group (case 1), we select SST, SSS, SSH, USSW, and VSSW as inputs. In addition to these parameters, we also include LON and LAT as input parameters in the second group (case 2). It can be clearly seen that the seven-parameter Ens-ML model (case 2) yields a significantly lower RMSE value than that of the five-parameter model (case 1) at all depths, and higher R^2 values (Figure 4). The vertically averaged RMSE and R^2 of the Ens-ML model with the seven-parameter model are $0.31\text{ }^\circ\text{C}$ and 0.89, respectively. While for the five-parameter model, the vertically averaged RMSE and R^2 are $0.69\text{ }^\circ\text{C}$ and 0.79, respectively. It should be noted that a significant difference in RMSE between the 7-parameter model and 5-parameter model was observed in the depth range of 250 to 600 m. For the 5-parameter model, the RMSE tends to increase from 250 to about 370 m, and then decrease with increasing depth. While for the 7-parameter model, the RMSE tends to decrease with increasing depth. This is an interesting question and requires further study. The results suggest that adding LON and LAT improves the estimation accuracy of the Ens-ML model in the SCS.

Table 4. Design of experiments.

Experiments	Training Methods
Case 1 (five parameters)	OSTS = Ensemble (SST, SSS, SSH, USSW, VSSW)
Case 2 (seven parameters)	OSTS = Ensemble (SST, SSS, SSH, USSW, VSSW, LON, LAT)

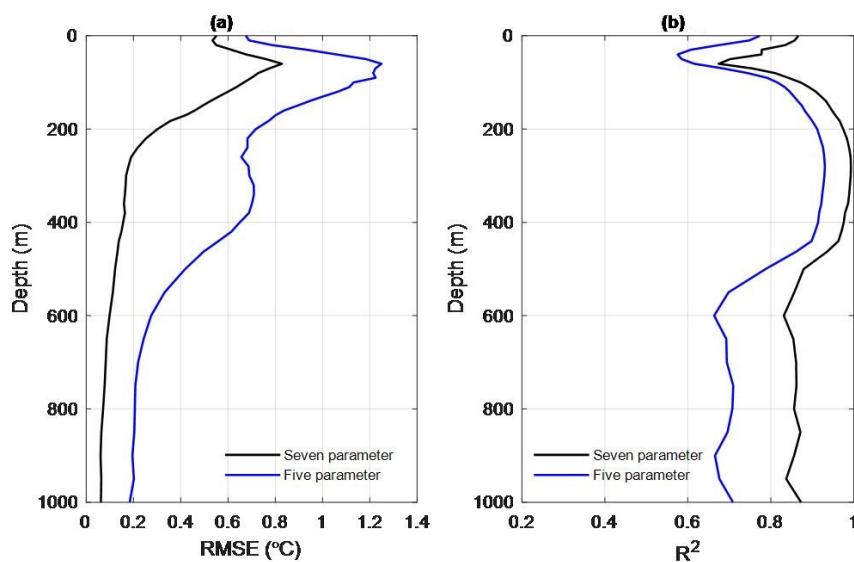


Figure 4. The estimation accuracy of the OSTs at different depths by the Ens-ML model based on (a) RMSE ($^{\circ}$ C) and (b) R^2 in different cases in 2019.

Next, we evaluate the performance of the seven-parameter Ens-ML model from different aspects in detail. First, some results of the Ens-ML-estimated OSTs at 50, 70, 100, 300, 500, and 1000 m depths in 2019 are shown in Figure 5. The Argo-derived OSTs at the same depths are used for evaluating the Ens-ML estimated OSTs and validate the estimated results by their differences, which are obtained from the Argo data minus Ens-ML model estimated data. We find good agreement between the Ens-ML model estimated OSTs and Argo OSTs maps at different depths. Most of the significant temperature features in the interior SCS can be effectively predicted from the sea surface data via the ensemble machine learning approach. For example, both datasets show that the temperature decreases with increasing depth. At 50 m depth, the temperature range is 22.5–27.1 $^{\circ}$ C, with a front trending from the southwest to the northeast separating the basin into two roughly equal water masses—cool water to the northwest, warm water to the southeast. The minimum temperature is found to the east of Vietnam about 111–113 $^{\circ}$ E, 13–16 $^{\circ}$ N, which coincides with the East Vietnam eddy [14]. The temperature difference between the Argo data and Ens-ML model estimation at 50 m depth only varies from 0.38 $^{\circ}$ C in the south to –0.21 $^{\circ}$ C in the north (Figure 5). At 70 m, the temperature varies from 25.0 $^{\circ}$ C in the northeast to 21.5 $^{\circ}$ C in the west of the SCS, which is also oriented northeast–southwest, with a positive temperature gradient toward southeastward. The spatial distribution of temperature difference between the Argo observation and model estimation at 70 m depth is similar to that of 50 m depth with a relatively large range from –0.43 $^{\circ}$ C to 0.40 $^{\circ}$ C. At 100 m depth, the temperature varies from 21.5 $^{\circ}$ C in the east to 18.5 $^{\circ}$ C in the west of the SCS with a positive temperature gradient toward southeast. The maximum difference is observed at about 112–116 $^{\circ}$ E, 11–14 $^{\circ}$ N, with a range of 0.21–0.32 $^{\circ}$ C. The results obtained show that the temperature tends to be stable with increasing depth. At 300 m depth, the temperature varies from 13.6 $^{\circ}$ C to 15.2 $^{\circ}$ C, while for the 500 and 1000 m depths, the temperature only varies from 7.6 $^{\circ}$ C to 8.6 $^{\circ}$ C and from 3.9 $^{\circ}$ C to 5.2 $^{\circ}$ C, respectively. Below 500 m depth, the temperature differences between the Argo-derived data and Ens-ML model estimation become relatively small with a range of –0.05–0.06 $^{\circ}$ C. Comparisons show that the maximum temperature differences between the Argo observation and model estimation occur at 50~70 m depth. Our results demonstrate that the proposed Ens-ML model presents a good performance in the estimation of OSTs in the SCS. A more detailed assessment of the Ens-ML model performance is given next.

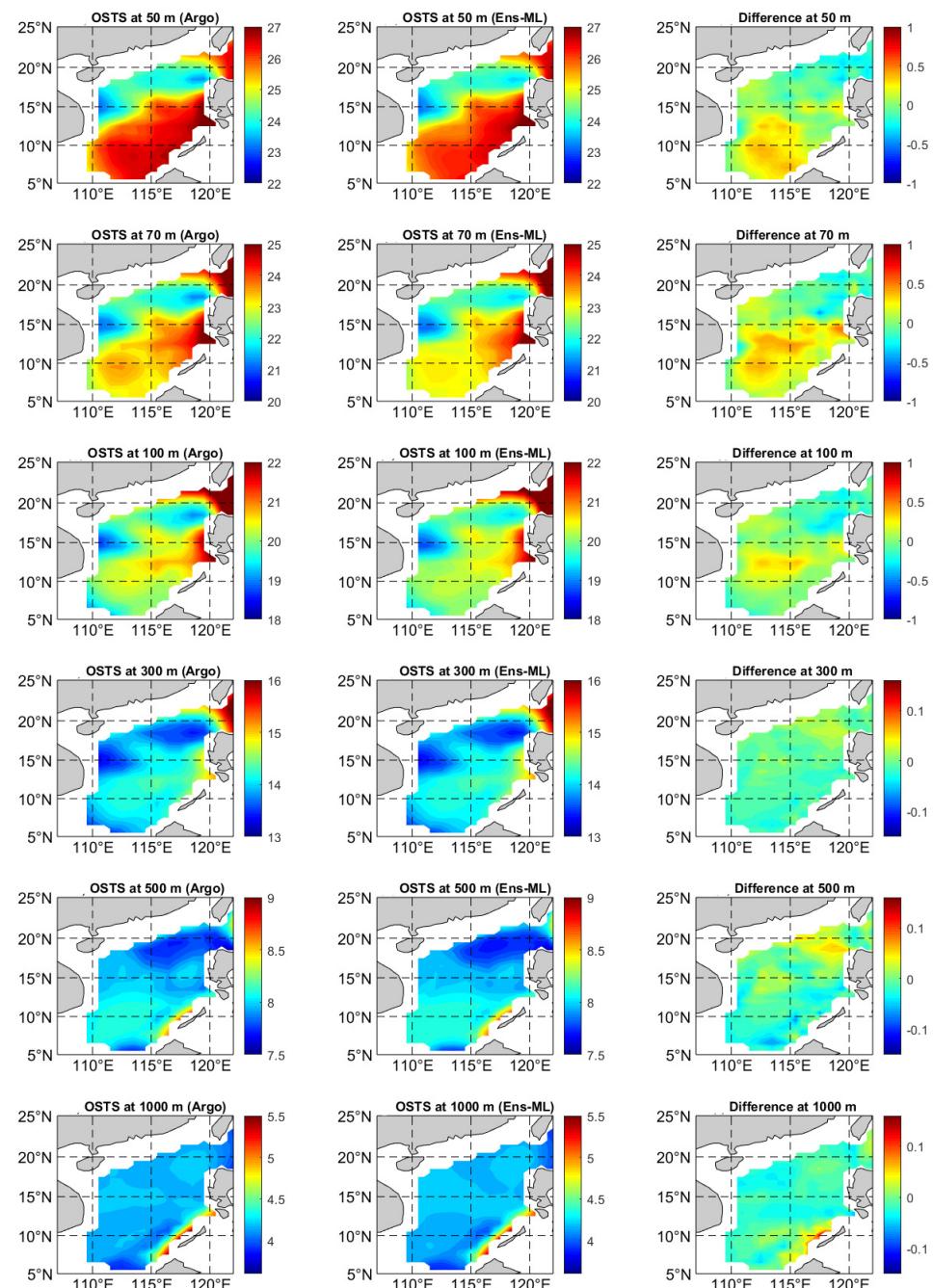


Figure 5. Argo-derived (left panel), Ens-ML estimated yearly mean temperature (middle panel) and the differences between the Argo and Ense-ML model (right panel) at different depths (50, 70, 100, 300, 500, 1000 m) in 2019.

Here, we quantitatively evaluate the accuracy of the OSTs estimation by the Ens-ML model using RMSE and R^2 as the performance measures (Table 5). The low RMSE and high R^2 mean more reasonable estimation accuracy. As shown in Table 5, the Ens-ML model shows good performances at all depths, although the performance measures vary with depth. Although the estimation accuracy shows some differences at different depths, the Ens-ML model is generally satisfactory. This also indicates that the Ens-ML model achieves a good performance for OSTs estimation in the SCS.

Table 5. Vertical distributions of RMSE ($^{\circ}\text{C}$) and R^2 for Ens-ML model at different depths.

Depth (m)	RMSE	R^2
2.5	0.51	0.86
10	0.49	0.85
20	0.51	0.83
30	0.57	0.77
50	0.72	0.72
70	0.73	0.75
100	0.62	0.87
150	0.45	0.94
200	0.26	0.97
300	0.12	0.99
400	0.11	0.97
500	0.08	0.87
600	0.05	0.83
700	0.04	0.86
800	0.03	0.85
900	0.02	0.85
1000	0.02	0.87

In addition, two sections (zonal section A along 18°N from 112°E to 118°E , and meridional section B along 113°E from 5.5°N to 18.5°N) were selected to further evaluate the vertical performance of the Ens-ML model (Figure 6). Zonal section A is an important section to investigate the influence of Kuroshio intrusion on the SCS [11]. Meridional section B, located between the tropics and subtropics, is an important section for studying meridional ocean transport in the SCS. Figure 7 shows the comparison of OSTs from Argo data and the Ens-ML model estimation along these two sections. It can be clearly seen that the spatial distribution of the Ens-ML model estimated OSTs exhibits good agreement with the Argo observations. Most of the main observed features of the vertical temperature structure are well reproduced by the Ens-ML model. In the upper 200 m, the temperature change with depth is large, varying from 27°C at the surface to 12°C at about 200 m. Below 200 m, the temperature only changes slightly varying from 10°C at 300 m to 4°C at 1000 m with the vertical gradient decreasing with depth.

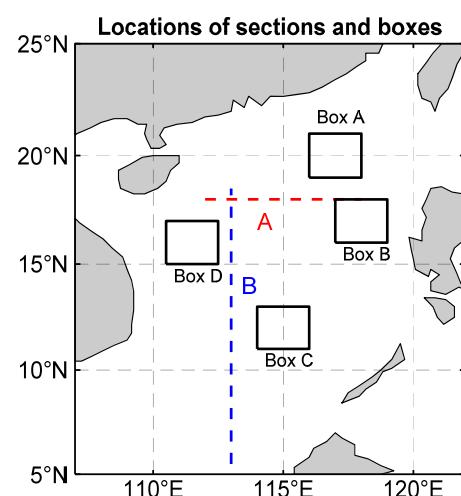


Figure 6. The zonal section A (latitude of 18°N), and meridional section B (longitude of 113°E) are defined to analyze the vertical variation of OSTs. Boxes a-d represent four boxes used in this study, Box A ($116\text{--}118^{\circ}\text{E}$, $19\text{--}21^{\circ}\text{N}$), Box B ($117\text{--}119^{\circ}\text{E}$, $16\text{--}18^{\circ}\text{N}$), Box C ($114\text{--}116^{\circ}\text{E}$, $11\text{--}13^{\circ}\text{N}$), and Box D ($110.5\text{--}112.5^{\circ}\text{E}$, $15\text{--}17^{\circ}\text{N}$).

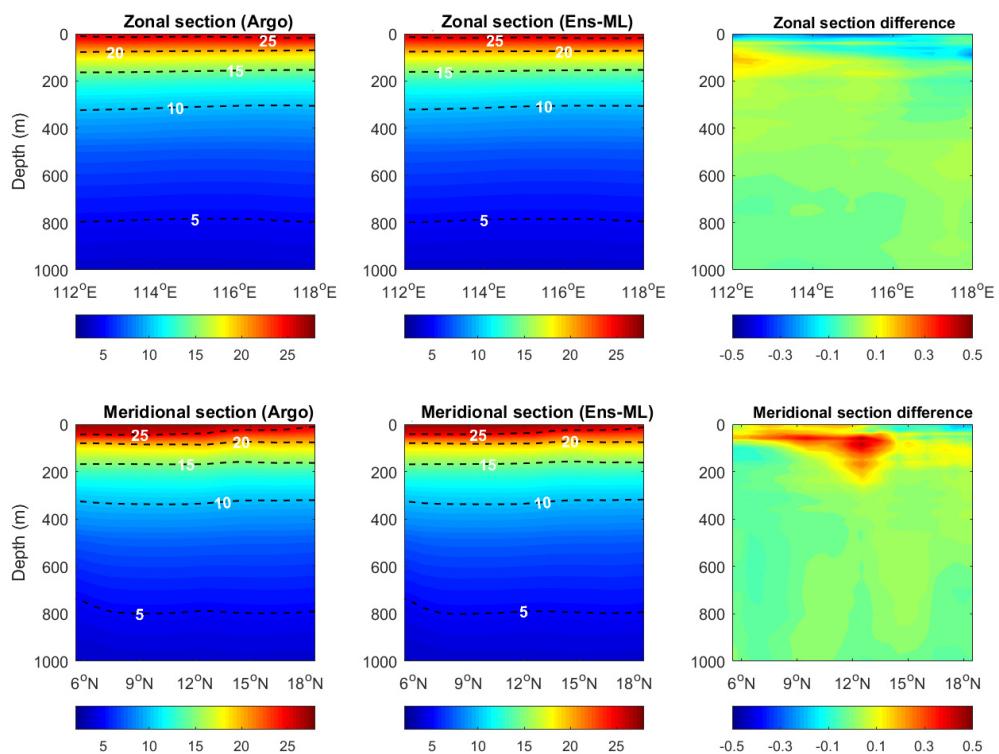


Figure 7. Argo-derived OSTs (left panel), Ens-ML model estimated OSTs (middle panel) and their differences (Argo minus Ens-ML, right panel) along the zonal and meridional sections in 2019.

For the zonal section (section A), in the upper 50 m, the estimated temperature value is slightly greater than the Argo value, with differences (Argo minus Ens-ML) up to -0.30°C . The maximum difference (exceeding 0.35°C) occurs between about 100 to 150 m, where the Ens-ML model-estimated temperature is relatively smaller than the Argo values. For the meridional section (section B), relatively large temperature differences (exceeding 0.40°C) are present in the depth range of 50 to 200 m between 6°N to 15°N , with Argo values higher than the estimated temperature values. Vertically averaged temperature errors for the zonal and meridional sections over 1000 m are about 0.05°C and 0.06°C , respectively. As the statistical analysis suggested, more than 98% of the section's points are within $\pm 0.10^{\circ}\text{C}$, and more than 99% of the section points are within $\pm 0.20^{\circ}\text{C}$ for the estimation.

A comparison between the Ens-ML model estimated temperature profiles and Argo profiles in 2019 are also presented in this study. Given the characteristics of bathymetry and Argo distributions, we selected four $2^{\circ} \times 2^{\circ}$ boxes designated A, B, C, and D in Figure 6 to further evaluate the performance of the Ens-ML model in the SCS. Box A ($116\text{--}118^{\circ}\text{E}$ and $19\text{--}21^{\circ}\text{N}$) is along the continental slope south of China. Box B ($117\text{--}119^{\circ}\text{E}$ and $16\text{--}18^{\circ}\text{N}$) coincides with the West Luzon eddy. Box C ($114\text{--}116^{\circ}\text{E}$ and $11\text{--}13^{\circ}\text{N}$) is in the central southern SCS. Box D ($110.5\text{--}112.5^{\circ}\text{E}$ and $15\text{--}17^{\circ}\text{N}$) is identical with the East Vietnam eddy. Figure 8 shows the temperature profiles in the four boxes from the Ens-ML model estimation and the Argo observations. We find a good agreement between the Ens-ML model-estimated profiles and Argo observed profiles. The RMSE between the averaged profiles from the Ens-ML model estimation and Argo observation are 0.17°C , 0.40°C , 0.61°C , and 0.53°C in boxes A, B, C, and D, respectively (Figure 8). Obviously, the Ens-ML model accurately reproduces the temperature profiles seen in the Argo observational data.

Scatter plots of temperature from the Argo observations and the Ens-ML model estimation show that most of the points are more or less evenly distributed around the line of equal value, giving a low RMSE and a high R^2 (Figure 9). These results also suggest that the estimated results of the Ens-ML model are reliable and performed well in the OSTs estimation in the upper 1000 m.

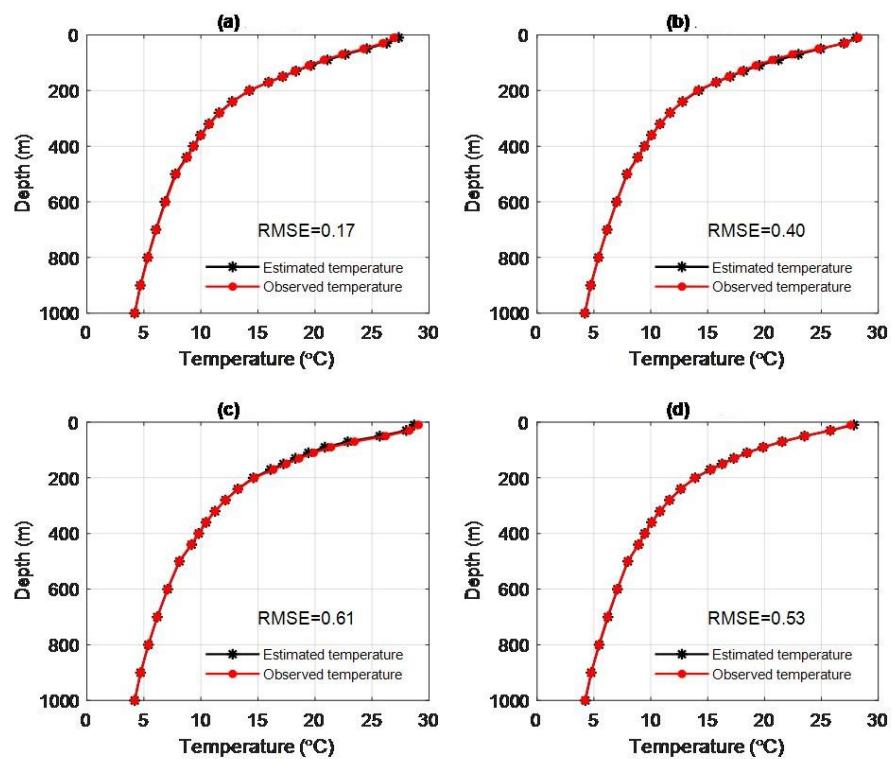


Figure 8. Comparison of area-averaged temperature profiles ((a–d) Boxes A–D) at different depths by the Ens-ML model estimation (black star line) and Argo observation (red dotted line).

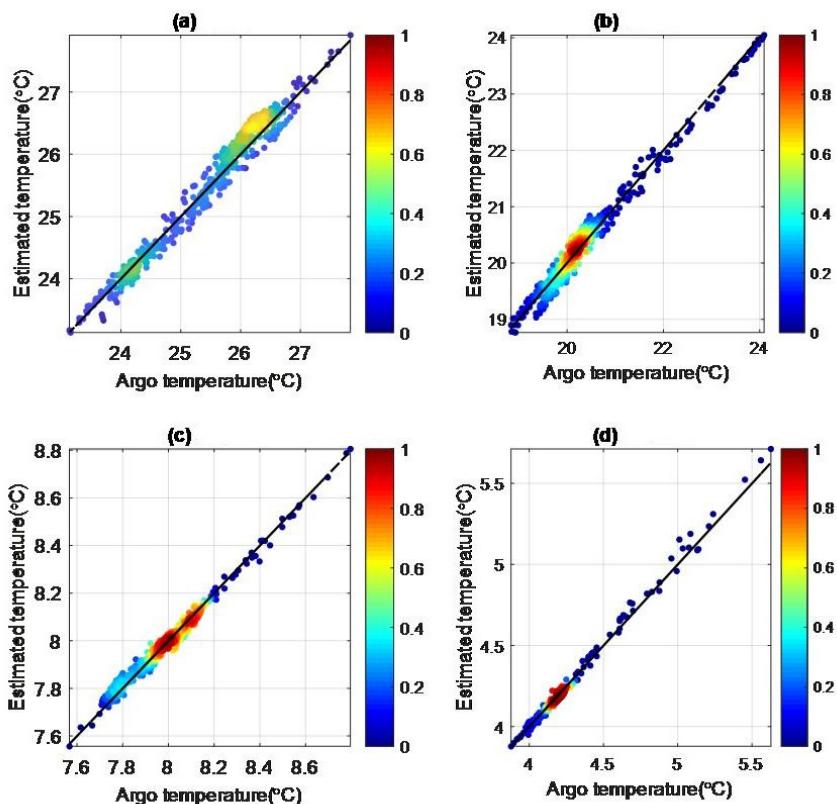


Figure 9. Scatter plots of temperature from the Ens-ML estimation and Argo observations at (a) 50 m, (b) 100 m, (c) 500 m, and (d) 1000 m in 2019. The data pairs are color-coded by data density at each temperature interval of $0.2\text{ }^{\circ}\text{C}$.

As discussed above, the Ens-ML model has been shown to have good performance in the yearly mean OSTs estimation over the SCS. A question then arises, how about its performance in different seasons. Next, we quantitatively assess the performance of the Ens-ML model for different seasons. February, May, August, and November of 2019 are selected to represent the winter, spring, summer, and fall seasons of the year. To improve the comparability of the model accuracy at different depths, we normalized the RMSE values (NRMSE is the RMSE divided by the standard deviation of the Argo temperature at that depth). Vertical distributions of the NRMSE and R^2 at different depths for different seasons are shown in Figure 10. On the whole, the Ens-ML performance is clear with relatively higher NRMSE and smaller R^2 in the upper 100 m. The NRMSEs in different seasons show first a downtrend and then an uptrend with a turning point occurring at about 200 or 300 m depths. The trend features of R^2 are also unstable and fluctuating, which show an uptrend from the surface to 300 m, and then a downtrend to the 1000 m depth. These suggest that the accuracy of the Ens-ML model generally decreases with increase of depth from 300 m. The reason for this variability may be that, in the deeper layer (below 300 m), the ocean phenomena have weaker surface manifestation, which are harder to interpret from sea surface information. In addition, relatively high NRMSE values in the upper 100 m might be related to the complex dynamic processes of the upper ocean.

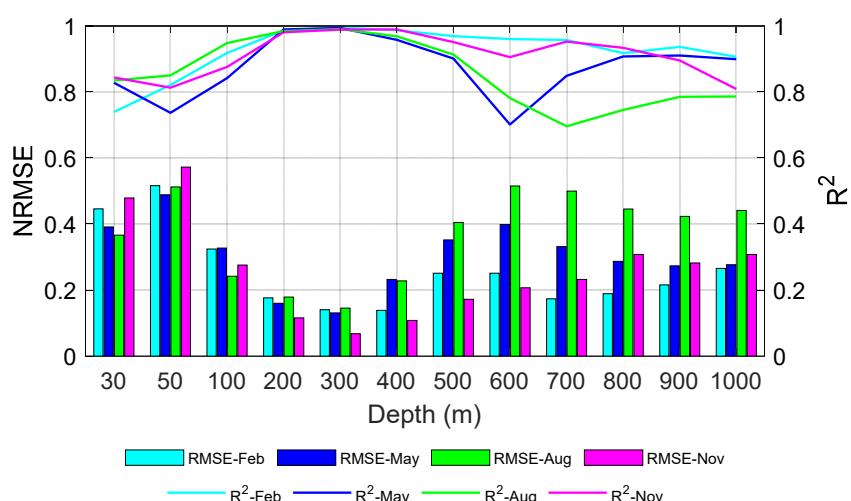


Figure 10. Seasonal performance of the Ens-ML model for ocean subsurface thermal structure estimation at different depths in the SCS in 2019. Cyan indicates February (winter), blue May (spring), green August (summer), and magenta November (autumn). Histograms display the NRMSE, and the lines display R^2 .

It is clearly seen that the prediction accuracy of the Ens-ML model varies with seasons. The vertical average NRMSE (R^2) values are $0.26\text{ }^\circ\text{C}$ (0.92), $0.30\text{ }^\circ\text{C}$ (0.88), $0.37\text{ }^\circ\text{C}$ (0.86), and $0.26\text{ }^\circ\text{C}$ (0.91) for February, May, August, and November, respectively. The highest accuracy appears in February, and the lowest accuracy appears in August. The OSTs in winter and autumn tends to be stable, and the magnitude of OSTs changes is small. The accuracy of the Ens-ML model is relatively high for all four seasons, indicating that the Ens-ML model has a good performance in the estimation of OSTs for different seasons.

4. Discussion and Conclusions

Ocean temperature has a large influence on marine environment and climate, with direct and indirect impacts on both natural and human activities, from the ecosystem system to the fishing industry. Hence, accurately estimating ocean subsurface thermal structures is of great importance for marine environmental monitoring, marine disaster prevention and climate changes. At present, machine learning models have been widely used to estimate ocean thermal structure. However, until now, most studies of OSTs estimation are

based on single models, not taking advantage of the potential improvement in estimation resulting from the use of an ensemble of models. To our knowledge, there have been no studies reported to use machine learning to estimate the OSTs in the SCS. Therefore, in this study, we introduce an ensemble machine learning model based on an ANN, which combines outputs from three individual models (XGBoost model, RandomForest model, and LightGBM model) to estimate the OSTs in the SCS.

The Ens-ML model was used to estimate the OSTs in the SCS with sea surface data as input and Argo observations for labeling. The performance of the Ens-ML model was examined in comparison with three individual models. Our results show that the Ens-ML model has the highest prediction accuracy; its average RMSE is 0.31°C , and its average R^2 is 0.89. The RMSE (R^2) values of the Ens-ML model are much smaller (larger) than those of the three individual models, indicating that the Ens-ML model performs better than the individual models. In addition to SSH, SST, SST, and SSW, geographic information (longitude and latitude) are two necessary parameters for accurately estimating the OSTs in the SCS, significantly improving the estimation accuracy of the Ens-ML model. Spatial distribution of the OSTs estimated by the Ens-ML model agrees well with the Argo observation, most of the observed OSTs features can be effectively detected from sea surface data via the Ens-ML model. The accuracy of the Ens-ML model for OSTs estimation generally decreases with increasing depth from 300 m, which is likely due to stable ocean stratification in the deeper ocean. The estimation accuracy of the Ens-ML model not only varies significantly with depth, but also varies with seasons. The vertical average NRMSE (R^2) for February, May, August, and November are 0.26°C (0.92), 0.30°C (0.88), 0.37°C (0.86), and 0.26°C (0.91), respectively. Obviously, the average NRMSE (R^2) in February is less (greater) than in May, August and November, suggesting that the estimation accuracy of the Ens-ML model in February is the best. Overall, the performance of the Ens-ML model is good for all four seasons, indicating that the Ens-ML model has a good seasonal applicability for the OSTs estimation in the SCS. This study demonstrates that a satellite-based ensemble machine learning approach to estimate OSTs can be accurate and reliable across different regions and times.

This study proposed and evaluated a multi-model ensemble approach to extend the satellite-derived oceanic data from the sea surface to subsurface. Our results show that this ensemble approach is robust and effective at estimating the OSTs in the SCS. However, it should be noted, the Ens-ML model, as a statistical method, enables us to estimate OSTs within the ranges of input parameters; this may lead to an underestimation of some extreme events. The reason for this is that the labeled OSTs of the Ens-ML model is in the range of input data for labeling. For any input data, the Ens-ML model can only estimate the OSTs that lies in the range of the input data used to train it. For those extreme events with OSTs exceeding the range of labeling data, the Ens-ML model may underestimate the temperature.

To extend this multi-model ensemble approach to the global scale, it is necessary to validate the model with different regions and data, and with better machine learning models. Besides different machine learning techniques, another important factor affecting estimation accuracy is the selection of input variables. It should be noted that the input variables used in this study might not be optimal combination for model. For example, previous studies have shown that wind stress, instead of wind speed, has a more significant impact on ocean temperature structure [62–64]. It would be interesting in future work to examine the different effects of wind stress and wind speed on OSTs estimation using a machine learning model. Once the OSTs has been accurately estimated, it can be used for practical applications, such as marine disaster prediction and acoustic propagation estimations [65]. Besides, application of the proposed multi-model ensemble approach to the estimation of other critical oceanic parameters, such as ocean subsurface salinity structure, oceanic density fields, and other oceanic variables, can be investigated as well. These are left for future study.

Author Contributions: Conceptualization, J.Q. and C.L.; methodology, J.Q., J.C. and D.L.; software, J.Q.; validation, L.G.; writing—original draft preparation, J.Q.; writing—review and editing, J.Q. and B.Y.; supervision, B.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 42176010, Natural Science Foundation of Shandong Province, China, grant number ZR2021MD022, Strategic Priority Research Program of the Chinese Academy of Sciences, grant number XDB42000000, National Natural Science Foundation of China, grant number 42006028, 42076022.

Data Availability Statement: The datasets presented in this study are publicly available.

Acknowledgments: The SST data was obtained from http://apdrc.soest.hawaii.edu/dods/public_data/NOAA_SST/OISST (accessed on 6 June 2021). The SSS data was obtained from https://data.catds.fr/cpdc/Ocean_products/ (accessed on 8 June 2021). The SSH data was produced and distributed by the Copernicus Marine and Environment Monitoring Service (CMEMS) (<http://www.marine.copernicus.eu>) (accessed on 12 June 2021). The SSW data was obtained from the <http://rda.ucar.edu/datasets/ds745.1/> (accessed on 16 June 2021). The Gridded Argo data product was obtained from <https://argo.ucsd.edu/data/argo-data-products/> (accessed on 17 June 2021).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Levitus, S.; Antonov, J.I.; Boyer, T.P.; Baranova, O.K.; Garcia, H.E.; Locarnini, R.A.; Mishonov, A.V.; Reagan, J.R.; Seidov, D.; Yarosh, E.S.; et al. World ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophys. Res. Lett.* **2012**, *39*, L10603. [[CrossRef](#)]
- Abraham, J.P.; Baringer, M.; Bindoff, N.; Boyer, T.; Cheng, L.; Church, J.A.; Conroy, J.L.; Domingues, C.M.; Fasullo, J.; Gilson, J.; et al. A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Rev. Geophys.* **2013**, *51*, 450–483. [[CrossRef](#)]
- Pearce, A.F.; Feng, M. The rise and fall of the “marine heat wave” off Western Australia during the summer of 2010/2011. *J. Mar. Syst.* **2013**, *111–112*, 139–156. [[CrossRef](#)]
- Oliver, E.C.J.; Donat, M.G.; Burrows, M.T.; Moore, P.J.; Smale, D.A.; Alexander, L.V.; Benthuysen, J.A.; Feng, M.; Gupta, A.S.; Hobday, A.J.; et al. Longer and more frequent marine heatwaves over the past century. *Nat. Commun.* **2018**, *9*, 1324. [[CrossRef](#)]
- Du, Y.; Zhang, Y.; Zhang, L.; Tozuka, T.; Ng, B.; Cai, W. Thermocline Warming Induced Extreme Indian Ocean Dipole in 2019. *Geophys. Res. Lett.* **2020**, *47*, e2020GL090079. [[CrossRef](#)]
- Wallace, J.M.; Rasmusson, E.M.; Mitchell, T.P.; Kousky, V.E.; Sarachik, E.S.; von Storch, H. On the structure and evolution of ENSO-related climate variability in the tropical Pacific: Lessons from TOGA. *J. Geophys. Res. Earth Surf.* **1998**, *103*, 14241–14259. [[CrossRef](#)]
- Planton, Y.Y.; Vialard, J.; Guilyardi, E.; Lengaigne, M.; McPhaden, M.J. The asymmetric influence of ocean heat content on ENSO predictability in the CNRM-CM5 coupled general circulation model. *J. Clim.* **2021**, *34*, 5775–5793. [[CrossRef](#)]
- Sprintall, J.; Tomczak, M. On the formation of central water and thermocline ventilation in the southern hemisphere. *Deep Sea Res. Part I: Oceanogr. Res. Pap.* **1993**, *40*, 827–848. [[CrossRef](#)]
- Qi, J.; Qu, T.; Yin, B.; Chi, J. Variability of the South Pacific Western Subtropical Mode Water and Its Relationship with ENSO During the Argo Period. *J. Geophys. Res. Oceans* **2020**, *125*, e2020JC016134. [[CrossRef](#)]
- Qu, T. Upper-layer circulation in the South China Sea. *J. Phys. Oceanogr.* **2000**, *30*, 1450–1460. [[CrossRef](#)]
- Qu, T.; Kim, Y.Y.; Yaremchuk, M.; Tozuka, T.; Ishida, A.; Yamagata, T. Can Luzon Strait transport play a role in conveying the impact of ENSO to the South China Sea? *J. Clim.* **2004**, *17*, 3644–3657. [[CrossRef](#)]
- Wang, G.; Xie, S.-P.; Qu, T.; Huang, R.X. Deep South China Sea circulation. *Geophys. Res. Lett.* **2011**, *38*, L05601. [[CrossRef](#)]
- Qi, J.; Du, Y.; Chi, J.; Yi, D.L.; Li, D.; Yin, B. Impacts of El Niño on the South China Sea surface salinity as seen from satellites. *Environ. Res. Lett.* **2022**, *17*, 054040. [[CrossRef](#)]
- Qu, T. Role of ocean dynamics in determining the mean seasonal cycle of the South China Sea surface temperature. *J. Geophys. Res. Earth Surf.* **2001**, *106*, 6943–6955. [[CrossRef](#)]
- Wang, A.; Du, Y.; Peng, S.; Liu, K.; Huang, R.X. Deep water characteristics and circulation in the South China Sea. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **2018**, *134*, 55–63. [[CrossRef](#)]
- Qu, T.; Song, Y.T.; Yamagata, T. An introduction to the South China Sea throughflow: Its dynamics, variability, and application for climate. *Dyn. Atmos. Oceans* **2009**, *47*, 3–14. [[CrossRef](#)]
- Wang, D.; Wang, Q.; Cai, S.; Shang, X.; Peng, S.; Shu, Y.; Xiao, J.; Xie, X.; Zhang, Z.; Liu, Z.; et al. Advances in research of the mid-deep South China Sea circulation. *Sci. China Earth Sci.* **2019**, *62*, 1992–2004. [[CrossRef](#)]
- Yao, Y.; Wang, C. Variations in Summer Marine Heatwaves in the South China Sea. *J. Geophys. Res. Oceans* **2021**, *126*, e2021JC017792. [[CrossRef](#)]

19. Chen, J.-M.; Li, T.; Shih, C.-F. Fall Persistence Barrier of Sea Surface Temperature in the South China Sea Associated with ENSO. *J. Clim.* **2007**, *20*, 158–172. [[CrossRef](#)]
20. Wu, C.-R.; Shaw, P.-T.; Chao, S.-Y. Assimilating altimetric data into a South China Sea model. *J. Geophys. Res. Earth Surf.* **1999**, *104*, 29987–30005. [[CrossRef](#)]
21. Xie, J.; Counillon, F.; Zhu, J.; Bertino, L. An eddy resolving tidal-driven model of the South China Sea assimilating along-track SLA data using the EnOI. *Ocean Sci.* **2011**, *7*, 609–627. [[CrossRef](#)]
22. Cornillon, P.; Stramma, L.; Price, J.F. Satellite measurements of sea surface cooling during hurricane Gloria. *Nature* **1987**, *326*, 373–375. [[CrossRef](#)]
23. Carnes, M.R.; Teague, W.J.; Mitchell, J.L. Inference of Subsurface Thermohaline Structure from Fields Measurable by Satellite. *J. Atmos. Ocean. Technol.* **1994**, *11*, 551–566. [[CrossRef](#)]
24. Ali, M.M.; Swain, D.; Weller, R.A. Estimation of ocean subsurface thermal structure from surface parameters: A neural network approach. *Geophys. Res. Lett.* **2004**, *31*, L20308. [[CrossRef](#)]
25. Buckingham, C.E.; Cornillon, P.C.; Schloesser, F.; Obenour, K.M. Global observations of quasi-zonal bands in microwave sea surface temperature. *J. Geophys. Res. Oceans* **2014**, *119*, 4840–4866. [[CrossRef](#)]
26. Su, H.; Li, W.; Yan, X.-H. Retrieving Temperature Anomaly in the Global Subsurface and Deeper Ocean from Satellite Observations. *J. Geophys. Res. Oceans* **2018**, *123*, 399–410. [[CrossRef](#)]
27. Mollo-Christensen, E.; Cornillon, P.; Mascarenhas, A.D.S. Method for Estimation of Ocean Current Velocity from Satellite Images. *Science* **1981**, *212*, 661–662. [[CrossRef](#)] [[PubMed](#)]
28. Chu, P.C.; Fan, C.; Liu, W.T. Determination of Vertical Thermal Structure from Sea Surface Temperature. *J. Atmospheric Ocean. Technol.* **2000**, *17*, 971–979. [[CrossRef](#)]
29. Cornillon, P.; Park, K.-A. Warm core ring velocities inferred from NSCAT. *Geophys. Res. Lett.* **2001**, *28*, 575–578. [[CrossRef](#)]
30. Osychny, V.; Cornillon, P. Properties of Rossby Waves in the North Atlantic Estimated from Satellite Data. *J. Phys. Oceanogr.* **2004**, *34*, 61–76. [[CrossRef](#)]
31. Cheng, H.; Sun, L.; Li, J. Neural Network Approach to Retrieving Ocean Subsurface Temperatures from Surface Parameters Observed by Satellites. *Water* **2021**, *13*, 388. [[CrossRef](#)]
32. Willis, J.K.; Roemmich, D.; Cornuelle, B. Combining altimetric height with broadscale profile data to estimate steric height, heat storage, subsurface temperature, and sea-surface temperature variability. *J. Geophys. Res. Earth Surf.* **2003**, *108*, 3292. [[CrossRef](#)]
33. Guinehut, S.; Dhomps, A.-L.; Larnicol, G.; Le Traon, P.-Y. High resolution 3-D temperature and salinity fields derived from in situ and satellite observations. *Ocean Sci.* **2012**, *8*, 845–857. [[CrossRef](#)]
34. Meijers, A.J.S.; Bindoff, N.; Rintoul, S. Estimating the Four-Dimensional Structure of the Southern Ocean Using Satellite Altimetry. *J. Atmospheric Ocean. Technol.* **2011**, *28*, 548–568. [[CrossRef](#)]
35. Wu, X.; Yan, X.-H.; Jo, Y.-H.; Liu, W.T. Estimation of Subsurface Temperature Anomaly in the North Atlantic Using a Self-Organizing Map Neural Network. *J. Atmospheric Ocean. Technol.* **2012**, *29*, 1675–1688. [[CrossRef](#)]
36. Charantonis, A.; Badran, F.; Thiria, S. Retrieving the evolution of vertical profiles of Chlorophyll-a from satellite observations using Hidden Markov Models and Self-Organizing Topological Maps. *Remote Sens. Environ.* **2015**, *163*, 229–239. [[CrossRef](#)]
37. Khedouri, E.; Szczechowski, C.; Cheney, R. Potential Oceanographic Applications of Satellite Altimetry for Inferring Subsurface Thermal Structure. In Proceedings of the OCEANS’83, San Francisco, CA, USA, 29 August–1 September 1983; pp. 274–280. [[CrossRef](#)]
38. DeWitt, P. Model decomposition of the monthly Gulf steam/Kuroshio temperature fields. *NOO Tech. Rep.* **1987**, 298.
39. Watts, D.R.; Sun, C.; Rintoul, S. A two-dimensional gravest empirical mode determined from hydrographic observations in the Subantarctic Front. *J. Phys. Oceanogr.* **2001**, *31*, 2186–2209. [[CrossRef](#)]
40. Su, H.; Huang, L.; Li, W.; Yang, X.; Yan, X. Retrieving Ocean Subsurface Temperature Using a Satellite-Based Geographically Weighted Regression Model. *J. Geophys. Res. Oceans* **2018**, *123*, 5180–5193. [[CrossRef](#)]
41. Yu, F.; Zhang, X.; Chen, X.; Chen, G. Altimetry-derived ocean thermal structure reconstruction for the Bay of Bengal cyclone season. *Ocean Dyn.* **2020**, *70*, 1449–1459. [[CrossRef](#)]
42. Prochaska, J.; Cornillon, P.; Reiman, D. Deep Learning of Sea Surface Temperature Patterns to Identify Ocean Extremes. *Remote Sens.* **2021**, *13*, 744. [[CrossRef](#)]
43. Chen, C.; Yang, K.; Ma, Y.; Wang, Y. Reconstructing the Subsurface Temperature Field by Using Sea Surface Data Through Self-Organizing Map Method. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1812–1816. [[CrossRef](#)]
44. Su, H.; Wu, X.; Yan, X.-H.; Kidwell, A. Estimation of subsurface temperature anomaly in the Indian Ocean during recent global surface warming hiatus from satellite measurements: A support vector machine approach. *Remote Sens. Environ.* **2015**, *160*, 63–71. [[CrossRef](#)]
45. Li, W.E.; Su, H.; Wang, X.; Yan, X. Estimation of global subsurface temperature anomaly based on multisource satellite observations. *J. Remote. Sens.* **2017**, *21*, 881–891.
46. Su, H.; Yang, X.; Lu, W.; Yan, X.-H. Estimating Subsurface Thermohaline Structure of the Global Ocean Using Surface Remote Sensing Observations. *Remote Sens.* **2019**, *11*, 1598. [[CrossRef](#)]
47. Su, H.; Wang, A.; Zhang, T.; Qin, T.; Du, X.; Yan, X.-H. Super-resolution of subsurface temperature field from remote sensing observations based on machine learning. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102440. [[CrossRef](#)]

48. Han, M.; Feng, Y.; Zhao, X.; Sun, C.; Hong, F.; Liu, C. A Convolutional Neural Network Using Surface Data to Predict Subsurface Temperatures in the Pacific Ocean. *IEEE Access* **2019**, *7*, 172816–172829. [[CrossRef](#)]
49. Lu, W.; Su, H.; Yang, X.; Yan, X.-H. Subsurface temperature estimation from remote sensing data using a clustering-neural network method. *Remote Sens. Environ.* **2019**, *229*, 213–222. [[CrossRef](#)]
50. Nardelli, B.B. A Deep Learning Network to Retrieve Ocean Hydrographic Profiles from Combined Satellite and In Situ Measurements. *Remote Sens.* **2020**, *12*, 3151. [[CrossRef](#)]
51. Meng, L.; Yan, C.; Zhuang, W.; Zhang, W.; Geng, X.; Yan, X.-H. Reconstructing High-Resolution Ocean Subsurface and Interior Temperature and Salinity Anomalies from Satellite Observations. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [[CrossRef](#)]
52. Xiao, C.; Chen, N.; Hu, C.; Wang, K.; Gong, J.; Chen, Z. Short and mid-term sea surface temperature prediction using time-series satellite data and LSTM-AdaBoost combination approach. *Remote Sens. Environ.* **2019**, *233*, 111358. [[CrossRef](#)]
53. Wolff, S.; O’Donncha, F.; Chen, B. Statistical and machine learning ensemble modelling to forecast sea surface temperature. *J. Mar. Syst.* **2020**, *208*, 103347. [[CrossRef](#)]
54. Gracia, S.; Olivito, J.; Resano, J.; Martin-Del-Brio, B.; de Alfonso, M.; Álvarez, E. Improving accuracy on wave height estimation through machine learning techniques. *Ocean Eng.* **2021**, *236*, 108699. [[CrossRef](#)]
55. Reynolds, R.W.; Rayner, N.A.; Smith, T.M.; Stokes, D.C.; Wang, W. An improved in situ and satellite SST analysis for climate. *J. Clim.* **2002**, *15*, 1609–1625. [[CrossRef](#)]
56. Boutin, J.; Vergely, J.; Marchand, S.; D’Amico, F.; Hasson, A.; Kolodziejczyk, N.; Reul, N.; Reverdin, G.; Vialard, J. New SMOS Sea Surface Salinity with reduced systematic errors and improved variability. *Remote Sens. Environ.* **2018**, *214*, 115–134. [[CrossRef](#)]
57. Hauser, D.; Tourain, C.; Hermozo, L.; Alraddawi, D.; Aouf, L.; Chapron, B.; Dalphinet, A.; Delaye, L.; Dalila, M.; Dormy, E.; et al. New Observations from the SWIM Radar On-Board CFOSAT: Instrument Validation and Ocean Wave Measurement Assessment. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5–26. [[CrossRef](#)]
58. Atlas, R.; Hoffman, R.; Ardizzone, J.; Leidner, S.M.; Jusem, J.C.; Smith, D.K.; Gombos, D. A Cross-calibrated, Multiplatform Ocean Surface Wind Velocity Product for Meteorological and Oceanographic Applications. *Bull. Am. Meteorol. Soc.* **2011**, *92*, 157–174. [[CrossRef](#)]
59. Roemmich, D.; Gilson, J. The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program. *Prog. Oceanogr.* **2009**, *82*, 81–100. [[CrossRef](#)]
60. Rajabi-Kiasari, S.; Hasanlou, M. An efficient model for the prediction of SMAP sea surface salinity using machine learning approaches in the Persian Gulf. *Int. J. Remote Sens.* **2019**, *41*, 3221–3242. [[CrossRef](#)]
61. Liu, Q.; Jia, Y.; Liu, P.; Wang, Q.; Chu, P.C. Seasonal and intraseasonal thermocline variability in the central south China Sea. *Geophys. Res. Lett.* **2001**, *28*, 4467–4470. [[CrossRef](#)]
62. Wang, Y.; Castelao, R.M.; Yuan, Y. Seasonal variability of alongshore winds and sea surface temperature fronts in Eastern Boundary Current Systems. *J. Geophys. Res. Oceans* **2015**, *120*, 2385–2400. [[CrossRef](#)]
63. Wang, Y.; Castelao, R.M. Variability in the coupling between sea surface temperature and wind stress in the global coastal ocean. *Cont. Shelf Res.* **2016**, *125*, 88–96. [[CrossRef](#)]
64. Wang, Y.; Liu, J.; Liu, H.; Lin, P.; Yuan, Y.; Chai, F. Seasonal and Interannual Variability in the Sea Surface Temperature Front in the Eastern Pacific Ocean. *J. Geophys. Res. Oceans* **2021**, *126*, e2020JC016356. [[CrossRef](#)]
65. Zheng, G.; Li, X.; Zhang, R.-H.; Liu, B. Purely satellite data-driven deep learning forecast of complicated tropical instability waves. *Sci. Adv.* **2020**, *6*, eaba1482. [[CrossRef](#)]