Parallel Prediction of Ocean Three-dimensional Fine Thermohaline Structure based on Surface Satellite Remote Sensing Data

Liming Yuan[1,2], Xingyue Du[1], Hao Jiang[1], Ming Zhao[1,2], Hanming Qian[1,2*], Baoqiang Zhang[1], Chunfei Lin[2]

[1]CSSC Ocean Exploration Technology Institute Co., Ltd., Wuxi, 214000, China

[2]Laboratory of Science and Technology on Marine Navigation and Control, China State Shipbuilding Corporation, Tianjin, 214000, China

*Corresponding Author.

Abstract:

In this study, using sea surface temperature and sea surface height data, the extreme gradient boosting (XGBoost) parallel model was selected through multi-model comparison to predict the three-dimensional temperature and salinity information. The 58 layers of global temperature and salinity information were forecasted within 1 minute, and the mean absolute error (MAE) was 0.319°C and 0.05psu, respectively. In particular, the prediction accuracy of the thermocline is poor, about 0.65°C, and the mid-deep layer is higher, about 0.3°C, which fully reflects the sensitivity of the model to the stratified structure of the ocean.

Keywords: Three-dimensional temperature and salinity, Satellite remote sensing data, XGBoost, Parallel, fine.

I. INTRODUCTION

The ocean occupies 71% of the earth's surface area and is one of the main factors regulating global climate development. As an important factor of ocean hydrological conditions, the ocean temperature and salinity structure is an important content of oceanography, and it also has an important impact on meteorology, navigation and combat.

In the early stage, the three-dimensional temperature and salinity information retrieval based on satellite remote sensing data mainly used dynamic methods [1-3] and statistical analysis methods[4-6]. The dynamic methods include: one is to use numerical models or dynamic constraints to establish a mapping between sea surface and underwater temperature and salinity information; the other is to invert the underwater three-dimensional temperature and salinity field based on the simplified dynamic model of the surface quasi-geostrophic equation. Among them, the most popular ocean numerical models include: MOM4 (Modular Ocean Mode 4.0), MITgcm (Massachusetts Institute of Technology Generalized Coordinate Model), NCOM (U.S. Navy Coastal Ocean Model), NLOM (U.S. Navy Coastal Ocean Model),

MICOM (Miami Equal Density Surface Coordinate Ocean Mode), HYCOM (Mixed Coordinate Ocean Mode), FVCOM (Limited Volume Nearshore Ocean Mode), POMgcs (Princeton University Generalized Coordinate Ocean Mode), ROMs (Regional Ocean Simulation System) and HANSOM ((Hamburg University shelf ocean model), etc., all of which can simulate three-dimensional changes in ocean temperature and salinity. Wang [7] and Liu et al. [8] used the "i-SQG" dynamic method to invert the temperature and salinity structure of the ocean and the subsurface density/flow field anomalies in a small area of the Gulf of Mexico, respectively.

Statistical methods are mainly divided into regression statistical methods, empirical orthogonal function decomposition and variational methods. Carnes et al. [9] used the least squares regression method to invert the temperature and salinity fields in the Northwest Pacific and Northwest Atlantic, and the inversion error was small. The regression analysis method adopted by Fox et al. [10] and Guinehut et al. [11] reconstructed the ocean three-dimensional temperature field with high temporal and spatial resolution; Maes [12] used the multivariate empirical orthogonal function-decomposition (m-EOF) method to obtain the coupled mode of temperature and salt changes, and further obtained the localized temperature and salt relationship; Fujii et al. [13] applied the vertical combined temperature and salt empirical orthogonal function EOF to the three-dimensional variational model, resulting in the error of the three-dimensional temperature field reconstructed from the sea surface temperature and height at about 1℃, and it can better reflect the El Niño and the La Niña process; Xiao X. J. et al. [14] developed an assimilation method based on the mutual constraint of sea surface height and sea surface temperature data under the three-dimensional variational framework (3DVAR), which effectively improved the estimation of ocean underwater temperature and salinity. The Ocean Variational analysis system OVALS (Ocean Variational analysis system) developed by the Institute of Atmospheric Physics, Chinese Academy of Sciences [15] can also effectively improve the simulation effect of underwater temperature and salinity fields.

The rapid development of artificial intelligence has also been widely used in the field of marine environment. It effectively improves the black box nature and expensive calculation cost of the dynamic method, and at the same time, it can appropriately avoid the disadvantages of statistical methods that are too dependent on long-term historical data and the chaotic nature of ocean development is difficult to capture. Wu et al. [16] used a self-organizing neural network model (SOM) based on sea surface temperature anomaly (SSTA) and sea surface height anomaly (SSHA) to invert the internal temperature structure of the North Atlantic Ocean; Su et al. proposed methods such as support vector machine (support vector machine) [17] and geographically weighted regression (GWR) [18] based on satellite sea surface data to reconstruct the underwater three-dimensional temperature field of the Indian Ocean; Yang X. et al. [19] used the random forest regression model to invert the ocean subsurface temperature field in different seasons based on SST, SSH, SSS, and SSW data; Ali et al. [20] used artificial neural networks to evaluate the temperature structure of the Arabian Sea from SST, SSH, and SSW data.

The previous researches focused on realizing the inversion of underwater temperature and salinity information, using data from a certain year during the research period to evaluate the accuracy of the model. In this way, higher inversion accuracy will be obtained, but the time series prediction of

temperature and salinity field is lacking, and the inversion accuracy of the internal structure of the ocean has not been explored in depth. However, unmanned submersible vehicle mission planning, safe and concealed navigation, target detection, etc. urgently need the support of future time sequence, fine and efficient three-dimensional temperature and salinity field information, and previous research cannot meet the needs of this actual operation. Based on the data of sea surface temperature and sea surface height, the XGBoost model is selected to use parallel computing to realize the high-efficiency and high-precision analysis and prediction of the future time series underwater three-dimensional temperature and salinity field through the comparative analysis of multiple models in different dimensions, and use the actual measurement data of Argo to evaluate the accuracy of the model, especially the accuracy assessment of the analysis and prediction of the internal stratification of the ocean. The XGBoost parallel model in this study focuses on the forecast of time series three-dimensional temperature and salinity information, and significantly improves the forecasting efficiency under the premise of ensuring accuracy, which can provide efficient and precise marine environment forecast information guarantee for offshore operating platforms.

II. STUDY AREA AND DATA

2.1 Study Area

Global Ocean prevalence of matter and energy transport and air-sea exchange, have a significant role in regulating climate change. The study area of this paper is the world (-89.75°~89.75°N, 0.25°~359.75°E).

2.2 Data

The surface satellite remote sensing data and Argo measured data used in this article are as follows:

The sea surface temperature data (SST) is from the Advanced Very High Resolution Radiometer (AVHRR) and the measured data using the optimal interpolation algorithm (optimum interpretation, OI fusion) to obtain the grid product. The time resolution is daily, the spatial resolution is 0.25°×0.25°.

The Sea Surface Height data (SSH) is multi-satellites (Jason-3, Sentinel-3A, HY-2A, Saral/AltiKa, Cryosat-2, Jason-2, Jason-1, T/P, ENVISAT, GFO, ERS1/2, etc.) merged grid delay product distributed by CMEMS. The time resolution is daily, the spatial resolution is 0.25°×0.25°.

Argo (Array for Real-time Geostrophic Oceanography) grid data contains subsurface temperature and salinity information, which is used for model training and prediction. The time resolution is monthly, the spatial resolution is 1°×1°, and the depth is 58 layers covering 0m~1975m.

III. METHOD

3.1 Data Preprocessing

Considering the time coverage of SST, SSH and Argo data, this study selects the data from 2004.01 to 2019.12 for analysis. In order to ensure the training efficiency and accuracy of the model, the data needs to be preprocessed.
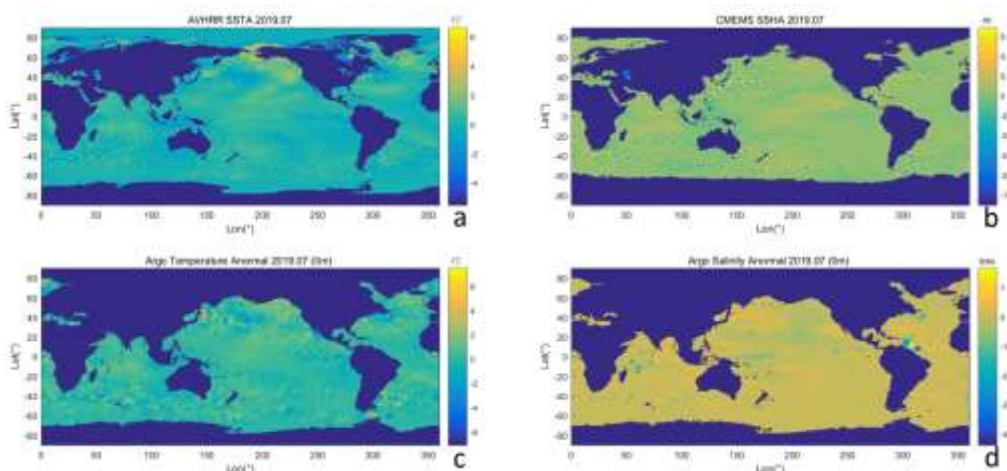

Fig 1: Distribution map of SSTA, SSHA, Argo-TA and Argo-SA in July 2019

The first is the unified processing of spatio-temporal resolution. The monthly average of SST and SSH data and the nearest spatial interpolation of Argo data are performed to obtain $0.25° \times 0.25°$ monthly SST, SSH and Argo data, respectively. The second is to remove the influence of climatic seasons. SST, SSH, Argo temperature and Argo salinity data respectively remove the average climatological field in different months from 2004.01 to 2019.12, and obtain the abnormal values of each parameter (SSTA, SSHA, Argo-TA, Argo-SA). Figure 1 shows the distribution of SSTA (Fig 1a), SSHA (Fig 1b), Argo-TA (Fig 1c) and Argo-SA (Fig 1d) in July 2019

3.2 Comparative Analysis of Different AI Models

In order to ensure the effectiveness and accuracy of the artificial intelligence model, this article first selects two sets of data for model comparison experiments. Use the 10m and 70m deep preprocessed data for 15 years from 2005.01 to 2018.12 as the training data, and the 10m and 70m depth data from 2019.01 to 2019.12 as the test data, and use the random forest regression model (RF) with 10 trees respectively. With XGBoost model training and testing, multi-dimensional comparison of accuracy, time, and hardware consumption is performed to facilitate the selection of a better model. The comparison results of the two models at the depth of 10m and 70m are shown in Table I and Table II, respectively. The XGBoost model is better than the RF model in terms of accuracy, time and hardware consumption under the same parameters. Therefore, this paper chooses the XGBoost model.

TABLE I. Comparison of 10m depth results between RF and XGBoost models

models	Training time /s	Model size /G	Trainingmemory /G	Testing time /s	MAE of temperature /°C	MAE of salinity /psu	Testing memory /G
RF(10trees)	~966	~66	~130	~388	~0.436	~0.137	~100
XGBoost(10trees)	~400	~0.00017	~20	~3.1	~0.375	~0.116	~0.78

TABLE II. Comparison of 70m depth results between RF and XGBoost models

models	Training time /s	Model size /G	Training memory /G	Testing time /s	MAE of temperature /°C	MAE of salinity /psu	Testing memory /G
RF(10trees)	~1000	~66	~130	~477	~0.702	~0.109	~100
XGBoost(10trees)	~380	~0.00017	~20	~3.5	~0.599	~0.096	~0.78

3.3 Model Training

The GBDT model is an iterative decision tree ensemble algorithm proposed by Jerome Friedman [21]. The extreme gradient boosting (XGBoost) is an optimized GBDT model, which has the advantages of fast calculation speed and high accuracy, and is widely used in ocean remote sensing. Based on the SST and SSH data, this study uses the XGBoost model to realize the analysis and forecast of the global three-dimensional temperature and salinity field. Figure 2 shows the specific process of the model. The model construction process is as follows three steps. First, the 15-year SSTA, SSHA, Argo-TA and Argo-SA data from January 2004 to December 2018 are randomly divided into independent training data sets (80%) and test data sets (20%), where SSTA and SSHA are used as model input variables, Argo-TA and Argo-SA are used as model training and testing labels. The second is to construct the XGBoost training model and continuously optimize the model through parameter tuning. The third is the prediction and accuracy evaluation of the XGBoost model based on SSTA and SSHA data from January to December 2019.
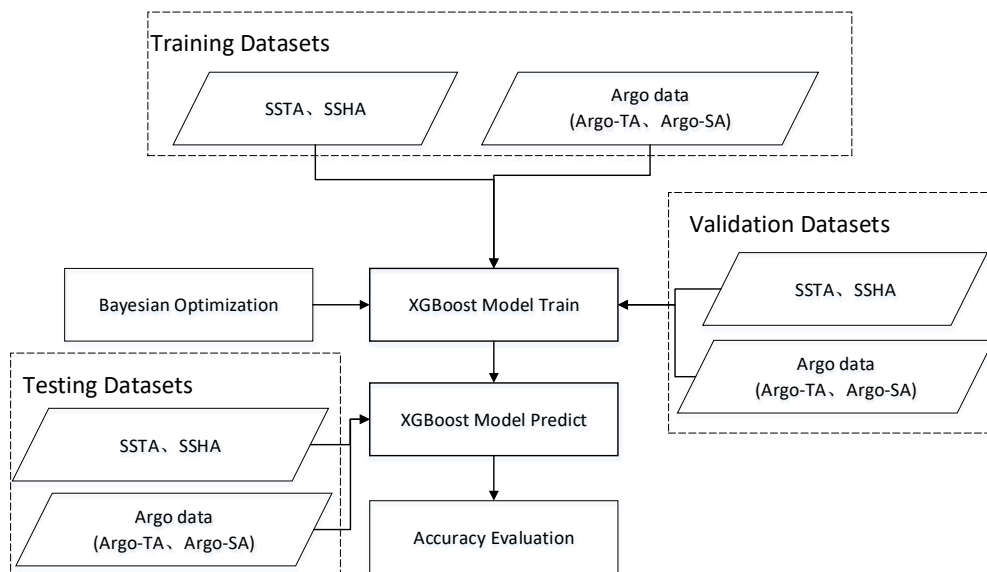
Fig 2: XGBoost model training and prediction process

Due to the significant differences in temperature and salinity characteristics at different depths, in order to ensure the accuracy of the model, this study constructed a temperature XGBoost model and a salinity XGBoost model for 58 different depths based on the training data. The temperature (salinity) XGBoost model constructed in this study can be calculated in 58-level parallel depth, using a high-performance server with 1 PowerEdge T640 MLK motherboard, Intel Xeon Silver 4210R 2.4G (2 x 10-core) CPU, 8×32G RDIMM 2933MT/s dual-rank memory, which fully guarantees the efficiency and accuracy of training and prediction models.

3.4 Model Evaluation

This study uses Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R^2 to evaluate the results of temperature (salinity) XGBoost training and prediction models. MAE reflects the actual forecast error magnitude, the smaller the MAE, the more accurate the forecast data; RMSE indicates the degree of deviation between the predicted value and the true value, the smaller the RMSE, the higher the prediction accuracy; R^2 characterizes the goodness of fit of the model, the larger the R^2, the better the goodness of fit of the model.

(1) MAE

$$MAE = \frac{\sum_{i=1}^{N} \left| Y_{obs,i} - Y_{model,i} \right|}{N} \tag{1}$$

Where N is the number of data, $Y_{obs,i}$ is the true value of Argo data, and $Y_{model,i}$ is the predicted value of the model.

(2) RMSE

$$RMSE = \sqrt{\sum_{i=1}^{N} \left(Y_{obs,i} - Y_{model,i} \right) \Big/ N}$$

(2)

Where N is the number of predictions.

(3) R^2

$$R^2 = \frac{\sum_{i=1}^{N} \left(Y_{model,i} - \overline{Y} \right)^2}{\sum_{i=1}^{N} \left(Y_{obs,i} - \overline{Y} \right)^2}$$

(3)

Where N is the number of predictions, $Y_{obs,i}$ is the true value of Argo data, $Y_{model,i}$ is the predicted value of the model, and $\overline{Y}$ is the true mean value of Argo data.

IV. RESULTS AND DISSCUSSION

4.1 Qualitative Analysis

In this study, 58 XGBoost temperature (salinity) prediction models at different depths were trained based on SST, SSH and Argo data from 2004 to 2018, and the model was used to analyze and forecast the three-dimensional temperature/salinity distribution in 2019. The temperature changes in winter (December, January, February) and summer (June, July, August) are more significant, so January and July are selected as representative to compare the results of the XGBoost temperature prediction model and Argo temperature at different depths so as to visually display the temperature prediction effect of the model.

Fig 3 and 4 show the spatial distribution of global ocean temperature based on the XGBoost temperature prediction model (left) and Argo data (right) at different depths of 0m, 100m, 500m and 100m in January (Fig 3) and July (Fig 4), respectively. Obviously, the global surface and underwater temperature distribution of the XGBoost temperature prediction model is consistent with the Argo temperature distribution at the same depth. It shows that the prediction results of the model can reflect the temperature distribution characteristics of different sea areas around the world. Secondly, the temperature characteristics predicted by the model at different depths are similar to those of Argo. Both showed that the temperature variation range and spatial distribution difference at depths of 0m and 100m were large, while the temperature variation range and spatial distribution difference below 500m decreased. This is because the thermal environment inside the ocean gradually stabilizes as the depth increases.
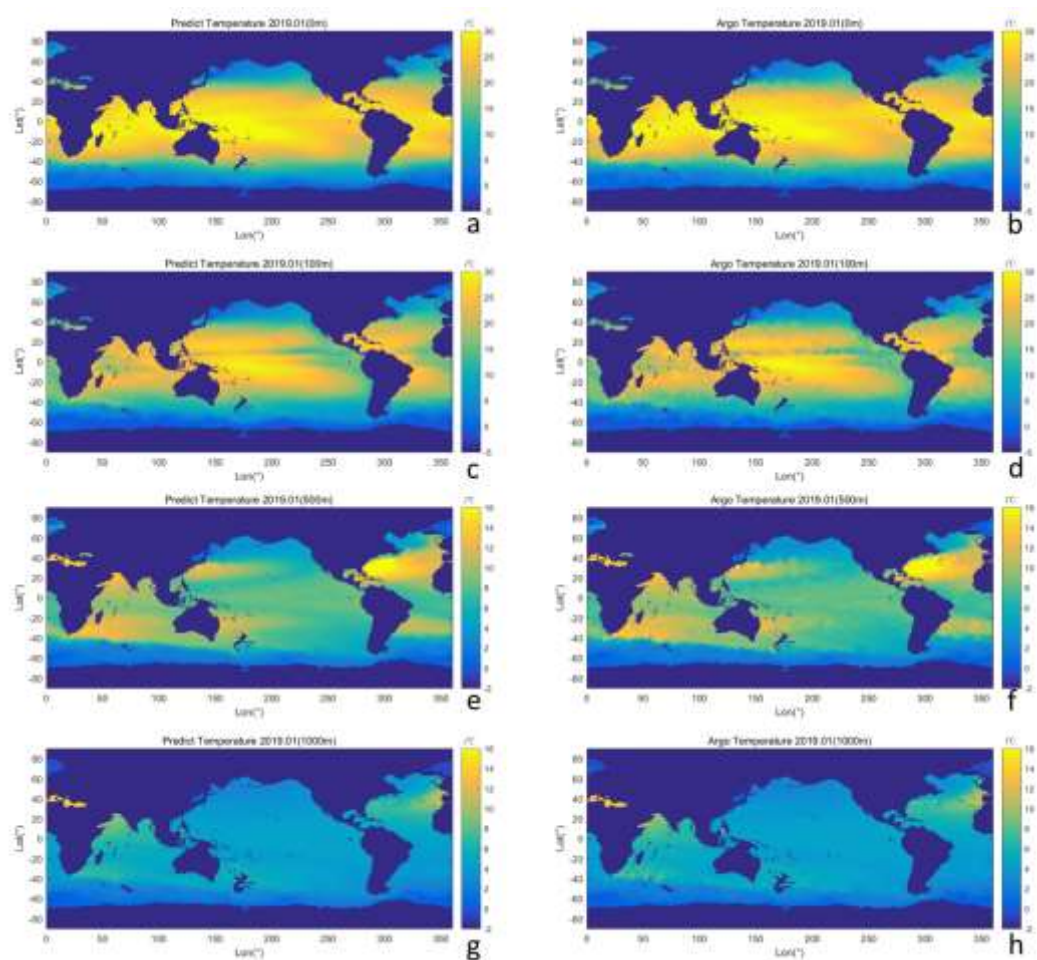
Fig 3: XGBoost prediction and Argo temperature distribution of 0m, 100m, 500m and 1000m in January 2019
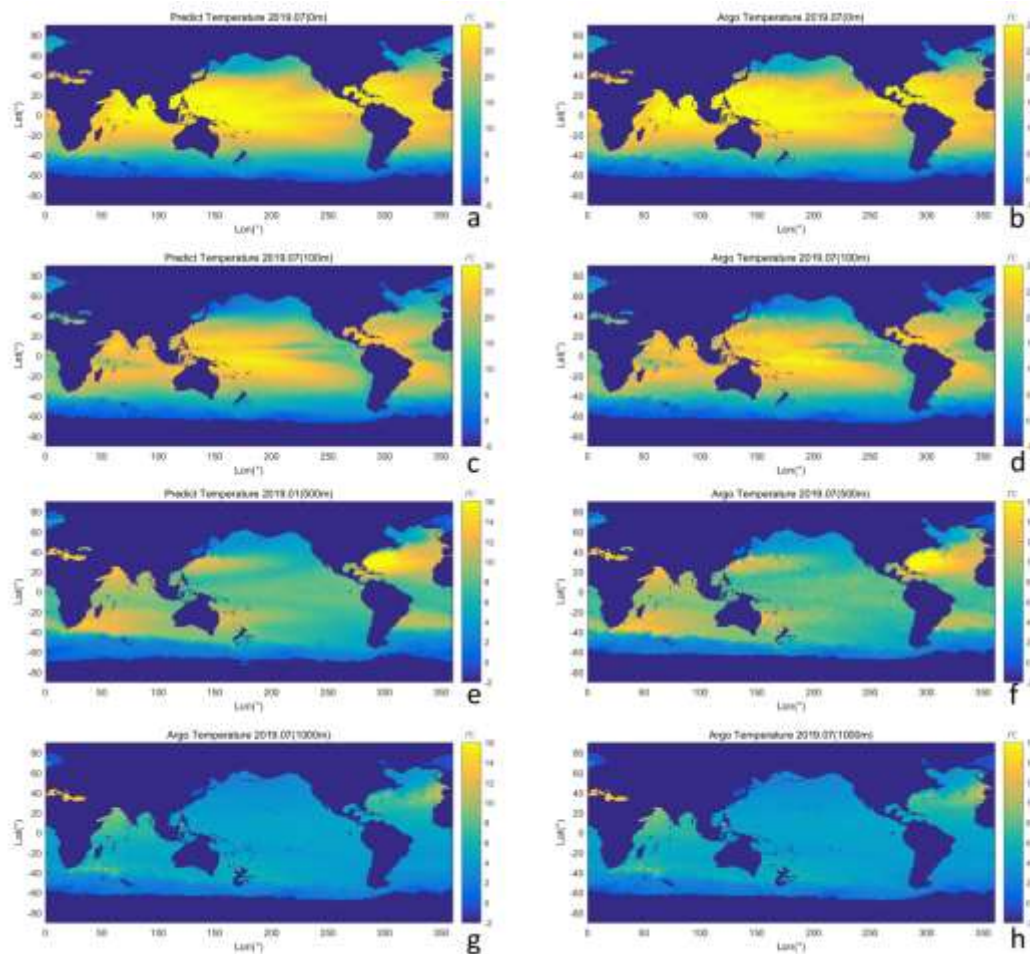
Fig 4: XGBoost prediction and Argo temperature distribution of 0m, 100m, 500m and 1000m in July 2019

4.2 Quantitative Analysis

The qualitative analysis shows that the temperature predicted by the XGBoost model established in this study is very consistent with the temperature distribution characteristics of the Argo data. We will quantitatively analyze the model through the model training efficiency and the MAE of temperature and salinity.

4.2.1 Model efficiency analysis

The XGBoost parallel model established in this study realizes the prediction of the three-dimensional temperature and salinity structure at different depths from January to December 2019. The average single-layer prediction time of the temperature model is 0.431s, the total prediction time of the 58-layer is 25.016s, and the MAE is within the range of 0.65℃; the average single-layer prediction time of the salinity model is 0.378s, the total prediction time of the 58-layer is 21.944s, and the MAE is in the range of 0.125psu. It can be seen that the XGBoost parallel model effectively improves model training and

prediction efficiency under the premise of ensuring the model accuracy.

4.2.2 Analysis of model temperature and salinity prediction accuracy

In order to evaluate the prediction accuracy of the XGBoost model in different stratified structures of ocean temperature and salinity, this paper analyzes the MAE of the temperature and salinity prediction for different months and different depths in 2019.

Fig 5a and 5b respectively show the winter (January, blue line), spring (April, orange line), summer (July, yellow line) and autumn (October, purple line) Temp-MAE and Salt-MAE curve with depth. It can be found that in the same season, Temp-MAE first increases and then decreases in the upper layer (above about 300m), and the middle and lower layer (below about 300m) shows a gradual decrease, with a peak at about 100m depth, mainly caused by the action of the ocean thermocline; in different seasons, the peak of Temp-MAE curve in spring and summer is slightly lower than that in autumn and winter, and the depth of the peak is slightly shallower than that in autumn and winter. This is consistent with the results of the thermocline distribution in winter and summer discovered by Zhou Y. X. and others [22]. The Salt-MAE curve shows a gradually decreasing trend with increasing depth, and there is no significant difference in different seasons.
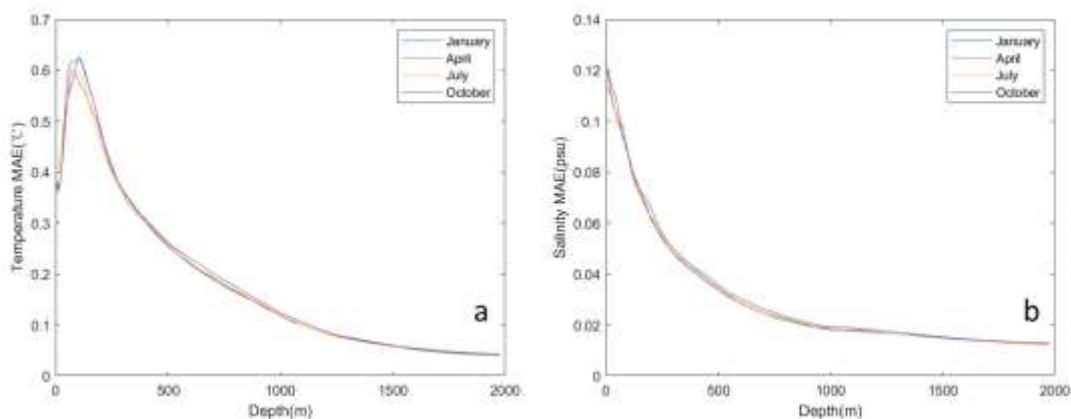

Fig 5: The Temp-MAE (a) and Salt-MAE (b) predicted by the XGBoost model at different depths in January, April, July, and October 2019

In order to further determine the sensitivity of the prediction model to the internal stratification structure of the ocean, this study analyzed Temp-MAE and Salt-MAE at different depths of 0m, 50m, 100m, 300m, 500m, 700m, and 1000m from January to December 2019, respectively as shown in Fig 6a and 6b. The surface, Temp-MAE is slightly higher in summer than other seasons and the regularity is not obvious, this is because the surface temperature is greatly affected by factors such as rainfall, wind and waves; at depths of 50m and 100m, Temp-MAE is slightly lower in spring and summer than in autumn and winter, consistent with the results in Fig. 6a, and also related to the distribution of the thermocline; below 300m, Temp-MAE changes basically steadily, and there is no obvious seasonal difference. In addition,

Temp-MAE has a larger value in the depth of 50m and 100m (within the range of 0.5~0.65°C), the distribution of Temp-MAE at depths of 0m and 300m is similar, and Temp-MAE below 500m gradually decreases with increasing depth. In general, the value of Salt-MAE is relatively small as shown in fig. 6b, the distribution of different months at the same depth is stable, and the same month gradually decreases with the increase of depth. It can be obtained that the XGBoost temperature model in this paper fully characterizes the influence of the thermocline, which can reflect the fine stratified structure of the ocean.
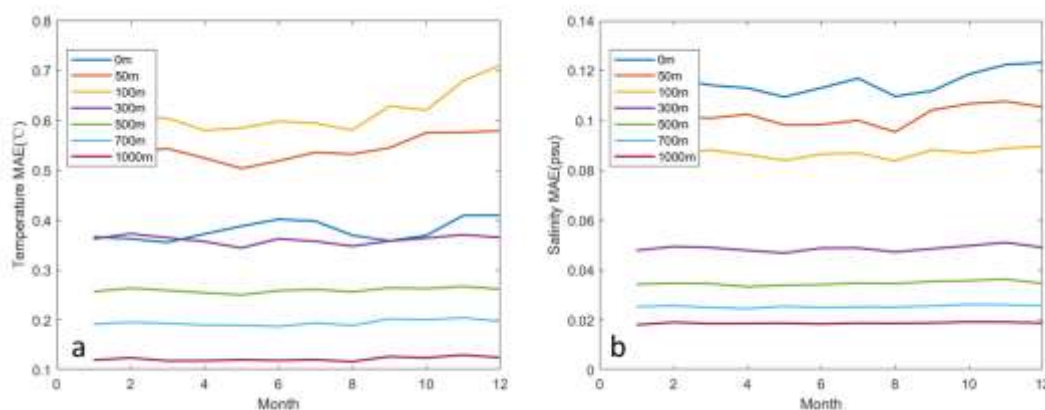

Fig 6: The Temp-MAE (a) and Salt-MAE (b) predicted by the XGBoost model in different months of 2019 at depths of 0m, 50m, 100m, 300m, 500m, 700m, and 1000m.

At the same time, we counted the Temp-MAE and Salt-MAE in winter, spring, summer and autumn of 58 levels from 0m to 1975m, as shown in Table 3. The average values of Temp-MAE in winter, spring, summer and autumn are 0.324°C, 0.310°C, 0.312°C and 0.328°C, respectively. The results also show that the temperature prediction accuracy in spring and summer is slightly higher than that in autumn and winter. The average values of Salt-MAE in winter, spring, summer and autumn are 0.052psu, 0.051psu, 0.051psu and 0.051psu, respectively. The salinity prediction accuracy of this model is high and there is no significant difference in different seasons.

TABLE III. The temperature and salinity MAE by the XGBoost model in different depths and different seasons in 2019

Depth /m	Temperature MAE /°C				salinity MAE /psu			
	Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn
0	0.380	0.372	0.390	0.380	0.119	0.112	0.113	0.118
5	0.393	0.378	0.392	0.393	0.122	0.114	0.117	0.122
10	0.373	0.368	0.385	0.375	0.120	0.113	0.115	0.119
20	0.390	0.382	0.421	0.391	0.115	0.109	0.110	0.114
30	0.436	0.420	0.471	0.445	0.110	0.107	0.105	0.112
40	0.502	0.471	0.506	0.514	0.107	0.103	0.102	0.110
50	0.550	0.524	0.530	0.566	0.103	0.101	0.098	0.106
60	0.581	0.567	0.554	0.608	0.100	0.097	0.096	0.102
70	0.600	0.588	0.575	0.633	0.099	0.094	0.093	0.098

80	0.623	0.600	0.583	0.645	0.096	0.092	0.091	0.095
90	0.637	0.596	0.590	0.647	0.092	0.089	0.088	0.092
100	0.649	0.590	0.592	0.643	0.088	0.086	0.086	0.088
110	0.645	0.579	0.585	0.635	0.084	0.082	0.083	0.084
120	0.630	0.570	0.576	0.628	0.080	0.078	0.079	0.081
130	0.615	0.561	0.565	0.612	0.078	0.076	0.075	0.078
140	0.599	0.551	0.552	0.597	0.076	0.073	0.073	0.076
150	0.580	0.538	0.536	0.581	0.075	0.071	0.072	0.074
160	0.561	0.526	0.521	0.564	0.073	0.070	0.070	0.073
170	0.544	0.515	0.508	0.546	0.070	0.068	0.068	0.071
180	0.525	0.505	0.492	0.528	0.068	0.066	0.067	0.070
200	0.488	0.474	0.458	0.489	0.064	0.063	0.063	0.066
220	0.455	0.441	0.431	0.453	0.060	0.059	0.060	0.062
240	0.428	0.414	0.407	0.427	0.057	0.056	0.056	0.058
260	0.403	0.390	0.388	0.401	0.054	0.053	0.053	0.055
280	0.383	0.372	0.370	0.381	0.051	0.050	0.051	0.052
300	0.367	0.356	0.357	0.365	0.049	0.048	0.048	0.050
320	0.354	0.342	0.345	0.353	0.047	0.046	0.047	0.048
340	0.342	0.330	0.334	0.340	0.045	0.045	0.045	0.047
360	0.329	0.320	0.323	0.328	0.044	0.043	0.044	0.045
380	0.318	0.311	0.314	0.319	0.043	0.042	0.043	0.044
400	0.309	0.302	0.305	0.310	0.042	0.041	0.041	0.042
420	0.298	0.293	0.295	0.302	0.040	0.039	0.040	0.041
440	0.288	0.283	0.285	0.292	0.039	0.038	0.039	0.040
460	0.279	0.274	0.276	0.282	0.037	0.037	0.037	0.038
500	0.261	0.255	0.259	0.265	0.035	0.034	0.035	0.036
550	0.240	0.235	0.238	0.248	0.032	0.031	0.032	0.033
600	0.225	0.219	0.219	0.233	0.030	0.029	0.029	0.030
650	0.209	0.205	0.204	0.218	0.027	0.027	0.027	0.028
700	0.195	0.191	0.190	0.203	0.026	0.025	0.025	0.026
750	0.182	0.177	0.178	0.190	0.024	0.023	0.024	0.025
800	0.171	0.165	0.167	0.179	0.023	0.022	0.023	0.023
850	0.158	0.153	0.155	0.165	0.021	0.021	0.021	0.022
900	0.147	0.142	0.143	0.152	0.020	0.020	0.020	0.021
950	0.134	0.130	0.131	0.139	0.019	0.019	0.019	0.020
1000	0.123	0.119	0.119	0.127	0.019	0.019	0.019	0.019
1050	0.111	0.109	0.109	0.118	0.018	0.018	0.018	0.019
1100	0.101	0.100	0.099	0.107	0.018	0.018	0.018	0.019
1150	0.093	0.092	0.091	0.098	0.018	0.018	0.017	0.018
1200	0.086	0.085	0.085	0.091	0.017	0.017	0.017	0.018
1250	0.081	0.079	0.079	0.083	0.017	0.017	0.017	0.017
1300	0.076	0.074	0.073	0.078	0.017	0.017	0.016	0.017
1400	0.068	0.065	0.065	0.069	0.016	0.016	0.016	0.016
1500	0.061	0.059	0.058	0.061	0.015	0.015	0.015	0.015
1600	0.055	0.053	0.053	0.055	0.014	0.014	0.014	0.015
1700	0.050	0.048	0.048	0.049	0.014	0.014	0.014	0.014
1800	0.046	0.044	0.044	0.046	0.013	0.013	0.013	0.013
1900	0.044	0.042	0.042	0.043	0.013	0.013	0.013	0.013
1975	0.043	0.041	0.041	0.041	0.013	0.013	0.013	0.013
Mean	0.324	0.310	0.312	0.328	0.052	0.051	0.051	0.053

V. CONCLUSION

This paper carried out a multi-dimensional comparison between the RF and the XGBoost model based on the SST and SSH data. The XGBoost parallel model was selected to analyze and forecast global ocean three-dimensional temperature and salinity information, combined with Argo temperature and salinity data to evaluate the model, including qualitative analysis of temperature prediction effectiveness and quantitative analysis of model efficiency and prediction accuracy of different depths and different seasons of temperature/salinity according to mean absolute errors.

The results show that: (1) The XGBoost parallel model established in this study can predict the 58-levels temperature and salinity from January to December 2019 within 30s, which significantly improves the efficiency of model prediction. (2) The XGBoost parallel model has a relatively high prediction accuracy for global three-dimensional temperature and salinity information, reaching the accuracy of the 58-layer temperature MAE of 0.319°C, and the salinity MAE of 0.054psu, where the temperature MAE of each layer is within the range of 0.65°C. (3) The spatial distribution characteristics of the XGBoost model predicted temperature are consistent with the Argo temperature.The difference in the spatial distribution and variation range of the predicted temperature at depths of 0m and 100m are larger, and the difference and variation range of the spatial distribution below 500m are reduced, which is related to the gradual stability as the depth increases. (4) The accuracy of temperature prediction in spring and summer is slightly higher than that in autumn and winter, while the prediction accuracy of the same season decreases first and then increases with changes in depth, especially in the vicinity of the thermocline, which is about 0.65°C, and the medium-deep layer is better at about 0.3°C.There is no obvious seasonal difference in the accuracy of salinity prediction, and it increases with the increase of depth. This is due to the influence of the stratified structure of the ocean and the distribution of the internal dynamic environment.

In summary, the XGBoost parallel temperature (salinity) prediction model constructed in this paper has significantly improved the prediction efficiency while maintaining high accuracy, and can reflect the distribution of ocean temperature stratification structure in detail, accurately and efficiently predict the three-dimensional temperature and salinity information f the global ocean, which can provide strong technical support for the marine environmental information guarantee urgently needed for offshore platform operations.

ACKNOWLEDGMENTS

appreciate the reviewers for their critical comments and suggestions to improve the original manuscript.

REFERENCES

[1] Lacasce J. H., Mahadevan A. Estimating subsurface horizontal and vertical velocities from sea-surface temperature, Journal of Marine Research, 2006, 64(5):695-721.

[2] Lapeyre G., Klein P. Dynamics of the Upper Oceanic Layers in Terms of Surface Quasigeostrophy Theory, J. phys. Oceanogr, 2006, 36(2):165-176.

[3] Isern-Fontanet J., Chapron B., Lapeyre G., et al. Potential use of microwave sea surface temperatures for the estimation of ocean currents, Geophysical Research Letters, 2006, 33(24):194-199.

[4] Hansen V. D., Thacker W. C. Estimation of salinity profiles in the upper ocean, J. Geophys. Res., 1999, 104(4):7921-7933.

[5] Agarwal N., Sharma R., Basu S., et al. Derivation of salinity profiles in the Indian Ocean from satellite surface observations, IEEE Trans. Geosci. Remote Sens., 2007, 4(2):322-325.

[6] Ballabrera-Poy J., Mourre B., Garcia-Ladona E., et al. Linear and non-linear T-S models for the eastern North Atlantic from Argo data:Role of surface salinity observations, Deep-Sea Research I, 2009, 56:1605-1614.

[7] Wang J., Flierl G. R., Lacasce J. H., Mcclean J. L., Mahadevan A. Reconstructing the ocean's interior from surface data, J. Phys. Oceanogr, 2013, 43, 1611-1626.

[8] Liu L., Peng S., Huang R. X. Reconstruction of ocean's interior from observed sea surface information, J. Geophys. Res. Oceans, 2017, 122, 1042-1056.

[9] Carnes M. R., Teague W. J., Mitchell J. L. Inference of subsurface thermohlaline structure from fields measurable by satellite, J Atmos Oceanic Technol, 1994, 11:551-566.

[10] Fox D. N., Teague W. J., Berron C. N., et al. The modular ocean data assimilation system, J. Atmos Oceanic Technol, 2002, 19:240-252.

[11] Guinehut S., Traon P., Larnicol G., et al. Combining Argo and remote-sensing data to estimate the ocean three dimensional temperature fields—a first approach based on simulated observations, Journal of Marine Systems, 2004, 46(1-4):85-98.

[12] Maes C. A note on the vertical scales of temperature and salinity and their signature in dynamic height in the western Pacific Ocean: Implications for data assimilation, Journal of Geophysical Research: Oceans (1978-2012), 1999, 104(C5):11037-11048.

[13] Fujii Y., Kamachi M. Three-dimensional analysis of temperature and salinity in the equatorial Pacific using a variational method with vertical coupled temperature-salinity EOF modes, Journal of Geophysical Research Atmospheres, 2003, 108(C9):65-66.

[14] Xiao X. J., Na H., Zhang Z. Q., et al. Variation assimilation using satellite data of sea surface temperature and altimeter, Journal of Tropical Oceanography, 2011(03):1-8. (Chinese)

[15] Zhu J., Zhou G., Yan C., et al. A three-dimensional variational ocean data assimilation system: Scheme and preliminary results, Science in China, 2006, 49(11):1212-1222. (Chinese)

[16] 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. Atmos. Ocean. Technol., 2012, 29, 1675–1688.

[17] 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.

[18] 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.

[19] Yang X., Su H., Li W. E., et al. Seasonal-spatial variations in satellite-derived global subsurface temperature anomalies, Journal of Remote Sensing, 2019, 23(5):997-1010. (Chinese)

[20] Ali M. M. Estimation of ocean subsurface thermal structure from surface parameters: A neural network approach, Geophysical Research Letters, 2004, 31(20):L20308.

[21] Friedman J. H. Greedy function approximation: A gradient boosting machine, Annals Stat., 2001, 29, 1189–1232.

[22] Zhou Y. X., Li B. L., Zhang Y. J., et al. Characteristics of the world ocean thermocline in winter and summer, Ocean Bulletin, 2002, 21(001):16-22.(Chinese)