RESEARCH   **Open Access**

Trend prediction of the 3D thermocline's lateral boundary based on the SVR method

Hongde Qin[1], Chong Wang[2,1], Yu Jiang[2,1], Zhongchao Deng[1*] and Wei Zhang[3]

Abstract

In recent years, with the use of the float-based wireless sensor network in the Argo project, a large amount of ocean data has been obtained. These data can be applied to analyze the oceanic thermocline, and the forecasting trend of the thermocline through the SVR method in machine learning is presented in this paper. Firstly, this paper refines the spatial resolution with the SVR method and determines the lateral boundary of the three-dimensional thermocline through the information entropy method. Combined with BOA Argo data from 2004 to 2015, this paper then predicts the thermocline trend (10°–25° S and 55°–80° E) over the next 4 years. The results show that the trend of the three-dimensional thermocline's lateral boundary can be effectively predicted with the application of SVR method.

1 Introduction

The ocean with abundant resources and broad development prospects has a vital significance to the humankind. With the increasing frequency of maritime and military activities, the rapid development of the marine economy, and the worsening marine environment, the marine science is increasingly raising public attention. The objective of marine scientific research are the composition, structure, property, distribution, genesis, and evolution rules of various natural phenomena relevant to the oceans, as well as the exploitation and utilization of marine resources. However, owing to the lack of oceanographic data, the analysis of the internal characteristics of the marine environment mainly focuses on large-scale regional and seasonal changes. To some extent, the implementation of Argo project has provided the data with more detailed and precise space and time intervals for marine environmental research. Meanwhile, many overseas and domestic scholars have utilized the data from Argo project to study the oceanic thermocline, circulation, water mass, and so forth [1–4]. Due to the limitation of data acquisition equipment and methods, more accurate data cannot be obtained at present. Over the past 15 years, the international Argo program has built a global ocean observing network of 3000 satellite-tracking automatic detection floats through the joint cooperation of more than 30 member countries in the world. It has extensively collected the global temperature and salinity data from the surface to 2000-m depth of ice-free deep ocean. As a kind of atypical wireless sensor network, the ocean observing network of Argo program is mainly applied for perceiving the relevant attributes of the ocean, such as temperature, salinity, depth, and other environmental information. Studying the Argo data is conducive to further exploring the internal state of the ocean, deeply analyzing how the oceans affect the global climate and obtaining innovative achievements in the fundamental researches and operational applications in the fields of marine, meteorological, fishery, and transportation.

The temperature and salinity of seawater are the basic elements of the oceans, and the variation of their spatial and time distribution is closely related to almost all the phenomena in the oceans. Therefore, the distribution and variation of the temperature and salinity will affect the movement of seawater and form different water masses by their different characteristics. In different sea areas, the spatial and temporal distribution of various marine environmental elements is extremely complicated, and the seawater (temperature, salt, density) spring layer is an important phenomenon in the ocean. Concerning the division of vertical boundaries of water masses, the existence and change of thermocline directly affects the mariculture and fishing, as well as the sound channel characteristics generated by the change of the acoustic cline, thereby affecting the submarine sonar communication system. The strong spring layer can also hinder the transport of

* Correspondence: dengzhongchao@hrbeu.edu.cn
[1]Science and Technology on Underwater Vehicle Technology, Harbin Engineering University, Harbin 150001, China
Full list of author information is available at the end of the article

nutrients between the upper and lower water layers, along with the vortex and convective heat exchange, play a natural "barrier" role. At the same time, the formation mechanism of thermocline is closely related to circulation, water mass, and internal wave of the ocean. Hence, the researches on thermocline are crucial to national defense, underwater communication, fishing, material diffusion, turbulent thermal diffusion, and other marine theoretical study. In recent years, with the rapid development of computer technology, a variety of data processing methods are generated. Regarding marine data with the characteristics of diverse types, large amounts, and complicated correlations, the traditional interpolation method can be replaced by the SVR method. As a statistical method, the traditional interpolation techniques require the model to re-run all the data. In terms of the dynamic SVR methods, it can only run on those new data, but not all the data again [5]. The predictive and generalization capacity of SVR depends on the choice of kernel function, while the traditional SVR method mainly chooses the kernel function based on the experience of certain risks in specific applications. Radial basis function (RBF) is an extensively used SVR kernel function with high prediction accuracy [6, 7].

In this paper, a method for predicting the trend of three-dimensional thermocline's lateral boundary is presented on the basis of SVR method. Specifically, the BOA Argo temperature data with spatial resolution of $1° \times 1°$ is initially refined. Then, the lateral boundary of the three-dimensional thermocline is determined with high-resolution data through the information entropy method, and its future variation trend is predicted as well.

The remainder of this paper is organized as follows. In Section 2, this paper introduces the source of Argo project data and the present research status of thermocline. In Section 5, the design and implementation of the algorithm are demonstrated in detail. In Section 4, the numerical and predicted results of the experiment are compared and analyzed. In Section 5, the summary and prospects of the paper are shown.

2 Related work

This section first introduces the details of the Argo project and proposes the shortcomings of ocean observations in the accurate determination of the thermocline. Then, the current situation and significance of the thermocline researches are given as well.

2.1 Argo project

The International Argo Program was launched in 2000 with the participation of over 30 countries and groups, including the USA, Japan, France, the UK, Germany, Australia, and China. Until December 2017, a total of 3891 active Argo profile buoys in the global oceans have been used to monitor seawater temperature, salinity, and currents (Fig. 1)
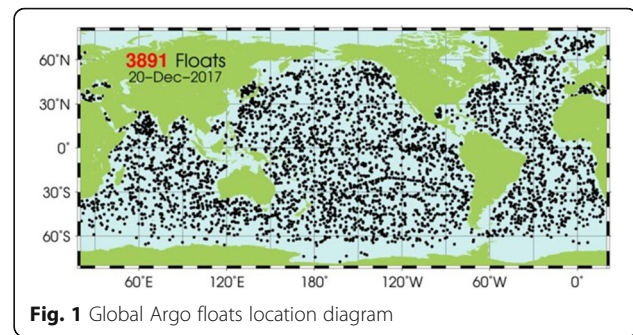
Fig. 1 Global Argo floats location diagram

(http://www.argo.ucsd.edu/, http://argo.jcommops.org/), which has basically achieved the construction goal of the Global Argo Observatory (maintaining 4000 buoys). The implementation of the Argo project can assist the researchers in accurately predicting such extreme weather or ocean events as typhoons and El Nino in the Pacific Ocean [8] [9]. Since the launch of the project, more than 12,000 floats have been deployed all over the world. In this condition, over 1.5 million temperature and salinity profiles have been obtained and are still increasing at a rapid rate every day. With the expansion of global Argo project to deep-sea Argo and biological Argo, the number of "core Argo" floats is expected to reach 4410 by 2020, and the sampling resolution of the upper ocean is raised as well.

Currently, the data provided by the Argo floats has been upgraded to version 3.1, which is utilized for further marine researches. Dong et al. adopt the temperature, salinity, and pressure profiles of the Argo floats to deduce the mixed-layer depth (MLD) of the Southern Ocean [10]. The estimation accuracy of Argo profiling float dataset for the temperature and heat storage in the upper North Atlantic Ocean is studied by Hadfield et al. as well [11]. The study of Guinehut et al. aims to analyze the contribution of the combination of high-resolution sea level and sea surface temperature satellite data with accurate but sparse in situ temperature profile data as given by Argo to the reconstruction of the large-scale, monthly mean, 200-m depth temperature fields [12]. Resnyanskii et al. use the Argo profiling floats dataset to estimate the means, variances, and three-dimensional spatial covariances of the temperature and salinity anomalies in the upper 1400 m ocean layer [13]. Maze et al. introduce how to conduct the unsupervised classification of Argo temperature profiles [14]. In recent years, it is difficult to deal with increasing volume of the marine data through the traditional mathematical statistics method, so the artificial intelligence method can be applied to process and analyze the massive data. However, the application range of the ocean observational data acquired through the conventional observation means or Argo floats is limited by some problems, such as inconsistent observation depth, discontinuous observation time, and spatial discrepancy.

The member countries of Argo project have analyzed the Argo data objectively and developed the gridded products [15–18]. As a supplement to the basic information on the global ocean phenomena, it greatly facilitates the further researches. Second Institute of Oceanography, SOA has also constructed a gridded dataset of Argo temperature and salinity in the global ocean through a more simple and effective objective analysis method, referring to the BOA Argo (http://www.argo.org.cn/).

The spatial range of BOA Argo data covers the global ocean (180°W–180°E, 79.5°S–79.5°N) with a spatial resolution of 1° × 1°. The seawater between 10 and 1950 m in depth is divided into 58 vertical standard layers, and the minimum distance between the two layers is 10 m. The gridded dataset could be used for studying the basic phenomena of physical ocean, but the precision of its data reaches the requirements when the thermocline is judged. In this paper, the SVR method will be applied to refine the data.

2.2 Research status and significance of thermocline

There exist the thermocline, halocline, pycnocline, and sound velocity spring layer in the ocean, and the thermocline refers to an area with a great change in the vertical gradient of seawater temperature. Since the eigenvalues of the thermocline mainly comprises the strength and thickness of the thermocline, and the depth of the upper boundary [19, 20], how to determine the three-dimensional boundary of thermocline and predict the variation trend of thermocline plays a key role in the analysis.

At present, a series of studies on the ocean temperature structure have been carried out. The researches on thermocline are meaningful for not only the theoretical study but also the national defense, underwater communications, and fisheries. In terms of fishing, the *Thunnus albacares* is one of the major targets of the oceanic tuna fishery worldwide, and it moves mostly inside the mixed layer and occasionally below the upper boundary of the thermocline, which is influenced by the temperature gradient greatly [21]. Meanwhile, there are many environmental factors affecting the fishing rate of the *Thunnus albacares*. Romena [22] pointed out that the distribution of adult *Thunnus albacares* was affected by the 20 °C isotherm, and Song [23] analyzed that the vertical distribution of *Thunnus albacares* was related to the thermocline. Concerning the underwater communication and military detection, the underwater acoustic communication is currently the unique means of communication in the ocean, but the acoustic propagation in the water is influenced by changes in temperature, salinity, and density. Therefore, a sudden change in seawater structure in the thermocline area will directly affect the sound transmission, leading to sonar failure [24]. In this condition, the researches

on the thermocline are significant to study the distribution of marine fishing grounds and underwater communication detection.

3 Theoretical method
3.1 Thermocline determination

In the analysis of thermocline, the thermocline should be initially determined, and the thermocline phenomena are usually described by the characteristic features of thermocline, namely the depth, the intensity, and the thickness. Hence, it is vital to determine the boundary of the thermocline and obtain the characteristic quantity of the thermocline. The traditional methods for determining the upper and lower boundaries of the thermocline are the vertical gradient method, the curvature extremum method, and the S-T method [25].

There exist some limitations of the traditional determination method for thermocline. Specifically, the vertical gradient method will cause the discontinuity between the two critical points of shallow water (less than 200 m in depth) and deep water (over 200 m in depth). After using the standard layer data to plot the temperature and depth curves, it is intuitive to determine the depth of the upper and lower bounds of the thermocline through the maximum curvature point method. In the case of insignificant curvature or multiple thermoclines, this method brings difficulties to data analysis. The S-T method is mainly applicable to the deep-water oceanic area, but not suitable for the areas obviously affected by solar radiation, precipitation, and diluted water. Since only the upper boundary of thermocline can be determined in the S-T method, this paper combines the "information entropy method" in machine learning with the traditional method for more precise determination [26]. The relevant principles and computational analysis process of the information entropy method are presented as below.

Information entropy: an indicator to measure the purity of the collective samples. Suppose that the proportion of the first k classes in the dataset is D, and p_k ($k = 1, 2, 3, …, |y|$) is the sample.

$$Ent(D) = -\sum_{k=1}^{|y|} p_k \log_2 p_k \qquad (1)$$

The entropy values increase as the uncertainty of the variables increases.

Information gain: the higher the information gain, the higher the purity acquired by performing the division through the attribute "a."

The calculation process in the information entropy method is shown as follows.

1. Select n samples attributes, then x_{ij} is the value of the ith sample's the jth attribute ($i = 1, 2, …, n; j = 1, 2, …, m$).

2. Normalize the index and make the homogeneity data a homogeneity.

Let

$$x_{ij} = |x_{ij}| \tag{2}$$

$$x_{ij} = \frac{x_{ij} - \min\{x_{ij}, ..., x_{nj}\}}{\max\{x_{1j}, jx_{nj}\} - \min\{x_{ij}, ..., x_{nj}\}} \tag{3}$$

then x_{ij} shall be the value of the ith sample's jth attribute ($i = 1, ..., n; j = 1, ..., m$).

3. Calculate the proportion of the ith sample in jth attribute.

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \quad (i = 1, ..., n; j = 1, ..., m) \tag{4}$$

4. Calculate the entropy value of the jth attribute.

$$e_j = -k \sum_{i=1}^{n} p_{ij} \ln\left(p_{ij}\right) \tag{5}$$

$$k = \frac{1}{\ln(n)} > 0 \ e_j \geq 0 \tag{6}$$

5. Calculate the information gain.

$$d_j = 1 - e_j \tag{7}$$

6. Calculate the weights of each index.

$$w_j = \frac{d_j}{\sum_{j=1}^{m} d_j} \tag{8}$$

7. Calculate the comprehensive score of each sample.

$$s_i = \sum_{j=1}^{m} w_j \cdot p_{ij} \tag{9}$$

Where s is the comprehensive score of each sample when it comes to form a thermocline; w is the important degree of each attribute for the formation of thermocline.

As the "information entropy method" combined with the traditional method can cover the shortage of only considering the strength, the thermocline can be more comprehensively and accurately determined, and then the lateral boundary of the three-dimensional thermocline can be determined.

3.2 SVR algorithm and prediction evaluation
3.2.1 Principles of support vector regression
We take the BOA Argo observational data as the training samples, and then train the model. Among the samples with the form of $\{(\mathbf{x}_1, z_1), ..., (\mathbf{x}_k, z_l)\}$, $\mathbf{x}_i \in R^n$ is the characteristic vector, and $z_i \in R^1$ is the target output. Considering that $C > 0$ and $\varepsilon > 0$, the standard form of support vector regression is represented as:

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i + C \sum_{i=1}^{l} \xi_i^* \tag{10}$$

subject to $\mathbf{w}^T \phi(\mathbf{x}_i) + b - z_i \leq \varepsilon + \xi_i,$

$\quad z_i - \mathbf{w}^T \phi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i^*,$

$\quad \xi_i, \xi_i^* \geq 0 \ , \ i = 1, ..., l$

Get the dual problem of SVR:

$$\min_{\alpha,\alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + \varepsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{l} z_i (\alpha_i - \alpha_i^*) \tag{11}$$

subject to $\mathbf{e}^T(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = 0$ (1)

$\quad 0 \leq \alpha_i, \ \alpha_i^* \leq C, \ i = 1, ..., l$

After solving problem (1), the approximate function of SVR is

$$f(x) = \sum_{i=1}^{l} \left(-x_i + x_i^*\right) K(x_i, x) + b \tag{12}$$

$Q_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) \equiv \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$ is the kernel function.

The selection for the kernel function affects the accuracy of SVR prediction significantly, and the BRF kernel function is relatively common and accurate at present.

$$\kappa(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{13}$$

3.2.2 Model evaluation
The precision, recall, and F1-measure are adopted to evaluate the prediction results of the algorithm, and TP, FP, TN, and FN in the in the formula are defined as follows.

TP, positive samples predicted as true; FP, positive samples predicted as false; TN, negative samples predicted as true; FN, negative samples predicted as false.

In terms of the prediction results, the precision ratio expresses the quantity of the true samples among all the positive samples. Among the positive samples, the positive category is possibly predicted to be positive, and the negative category is possibly mispredicted to be positive.

$$\text{Precision} \ = \ \frac{TP}{TP \ + \ FP} \tag{14}$$

Aiming at the original samples, the recall rate represents the quantity of the correctly predicted samples among all the positive samples. Among the positive samples, the original positive category is probably predicted to be positive, and the original negative category is probably mispredicted to be positive.

$$\text{Recall} \ = \ \frac{TP}{TP \ + \ FN} \tag{15}$$

Regarding the evaluation results, the precision ratio and the recall rate are expected to be as high as possible. However, they are contradictory in some cases; for instance, the higher precision ratio is frequently accompanied by recall rate. In this circumstance, the F1-measure can be applied for a balanced assessment.

$$F1 \ = \ \frac{2 \cdot TP}{2 \cdot TP \ + \ FP \ + \ TN} \tag{16}$$

4 Experimental verification and result analysis

In this section, the experimental method is validated, and the lateral boundary of thermocline is determined by combining this method with the information entropy method. The variation trend of the lateral boundary of thermocline is predicted as well. The experiment adopts Ubuntu Linux 16.04, Python 3.6, scikit-learn 1.9.2 as experimental test platform.

4.1 Data selection

This paper utilizes the data from the Global Ocean Argo Gridded Dataset (BOA Argo) in the period of 2004–2016. The global marine temperature data (79.5°S–79.5°N, 180°W–180°E) from January 2004 to December 2016

Table 1 The evaluation for prediction results

Year-Mon	Tp	Ture	Predict	Precision	Recall	F1
1601	122	200	156	0.61	0.782051282	0.685393258
1602	124	167	142	0.74251497	0.873239437	0.802588997
1603	54	132	71	0.409090909	0.76056338	0.532019704
1604	54	153	65	0.352941176	0.830769231	0.495412844
1605	39	127	42	0.307086614	0.928571429	0.461538462
1606	66	117	84	0.564102564	0.785714286	0.656716418
1607	81	122	110	0.663934426	0.736363636	0.698275862
1608	87	165	107	0.527272727	0.813084112	0.639705882
1609	81	159	98	0.509433962	0.826530612	0.630350195
1610	95	176	118	0.539772727	0.805084746	0.646258503
1611	133	215	159	0.618604651	0.836477987	0.711229947
1612	121	208	134	0.581730769	0.902985075	0.707602339

is selected from China Argo real-time data center (http://www.argo.org.cn/). A large number of experimental results demonstrate that the thermocline exists in the seawater with a depth less than 500 m. Hence, we choose the thermocline in shallow seawater with a depth less than 500 m for analysis. Specifically, this paper selects the sea area (10°–25° S and 55°–80° E) for study [27], which is illustrated in the gridded area in Fig. 2.

4.2 Method validation

In order to predict the future trend of the three-dimensional thermocline's lateral boundary, this paper adopts the SVR method in machine learning on the basis of the original data from 2004 to 2015 to conduct a prediction of 2016. Meanwhile, the precision, recall, and F1-measure are used to evaluate the results of the algorithm as shown in Table 1.

As seen from the above table, the accuracy rate can reach above 0.5, especially in the winter and summer
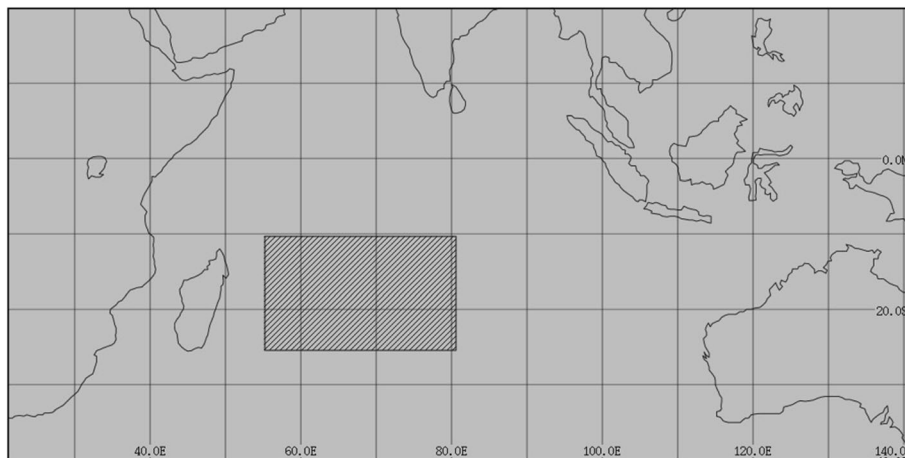
Fig. 2 The study area of thermocline

Fig. 3 The flow diagram of high-resolution data refinement
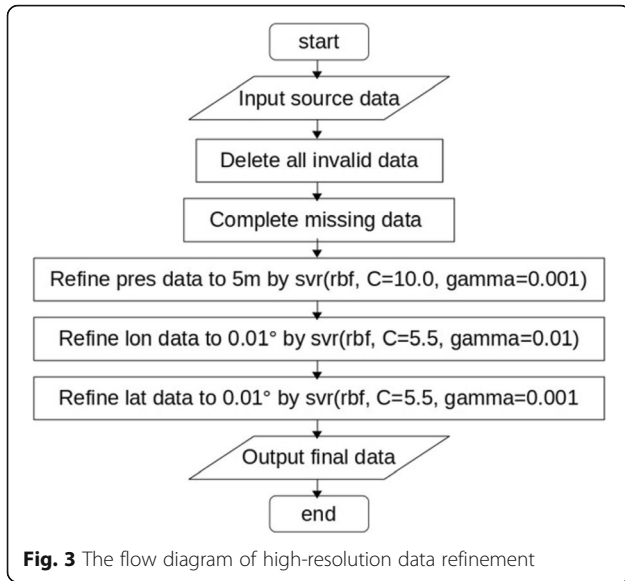
when the thermocline appears obvious, and both the recall rate and F1 value obtain higher results. Therefore, it can be concluded that the SVR method used in this paper can accurately predict the variation trend of thermocline's lateral boundary.

4.3 Data preprocessing
4.3.1 Data refinement
The high-resolution marine temperature and depth data for more accurate thermocline determination and trend prediction, but the fitness of the gridded data from BOA Argo is far below our requirements at present. This paper uses the SVR method to refine the BOA Argo data to eventually obtain high-resolution data of $0.01° \times 0.01° \times 5$ m.

When adopting the SVR method for data refinement, the correct values should be set. As the values differ in the longitude, latitude, and depth, we need to refine the longitude, latitude, and depth data separately through the process displayed in Fig. 3.

According to the above process, the high-resolution oceanographic data can be acquired in the SVR method. Taking the February 2016 data with obvious temperature changes as an example, the temperature distribution in the study area under real data is compared with that under the high-resolution data as shown in Fig. 4.

Figure 4 demonstrates the temperature distribution of the real data on the left and the temperature distribution of the high-resolution data on the right. The overall trend of the temperature distribution under the real data consists with that under the high-resolution data, but the temperature distribution under the high-resolution data is more refined. Specifically, the regional boundary of temperature change is more obvious, the area with the temperature jump in the upper right part of the figure shows more prominent temperature gradients, and the excessive change of boundary temperature is more precise. In this condition, the high-resolution data after refinement with a great research value is easier to observe and analyze.

4.3.2 Determination of thermocline
The thermocline can be more accurately determined by the high-resolution data after refinement, and we introduce a concept of temperature strength to judge the thermocline [28].

$$l = \frac{t_{n+1} - t_{n-1}}{d_{n+1} - d_{n-1}} \quad n = 2, 3, 4, \ldots \tag{17}$$

Where l is the temperature strength, t is the temperature, d is the depth, and n is the layer number.

Fig. 4 The comparison of the temperature distribution in the study area under the real data and high-resolution data
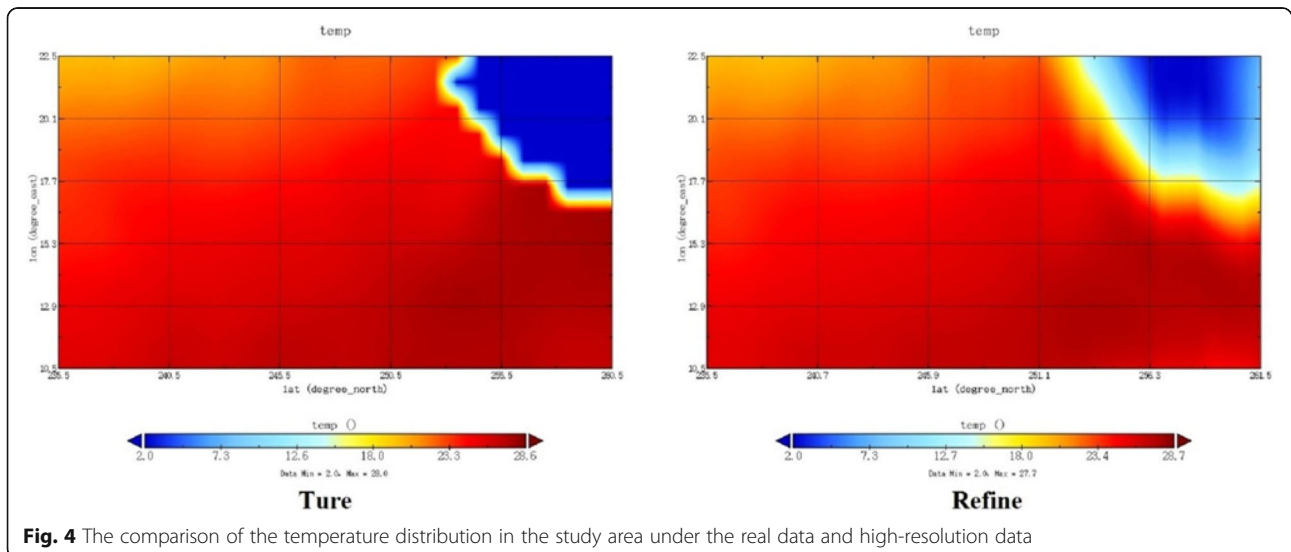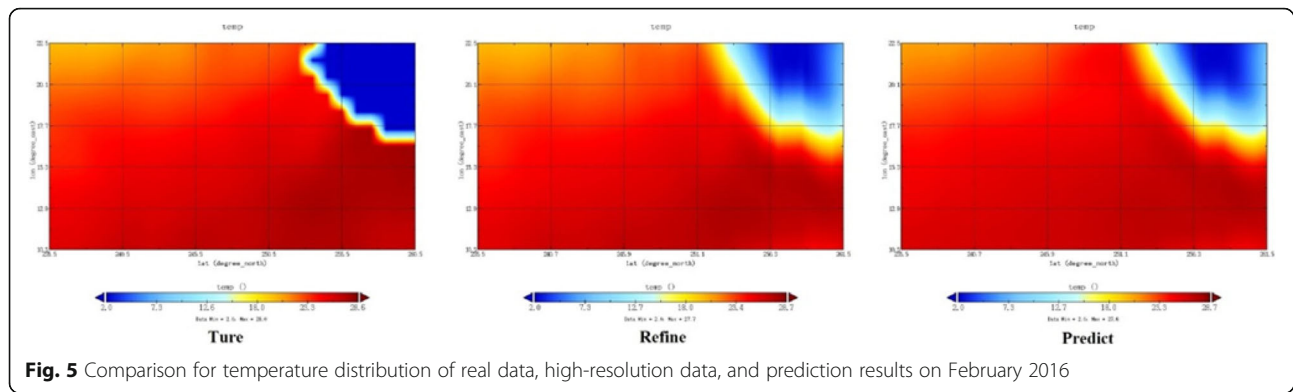
Fig. 5 Comparison for temperature distribution of real data, high-resolution data, and prediction results on February 2016

It is assumed that the thermocline does not exist on the surface, and the test results are supposed not to be affected irrespective of the first layer.

According to Jiang's research, the possible coexistence of the thermocline and reverse thermocline is observed, so the temperature strength can be positive or negative. To simplify calculation, here we make

$$l = |l| > 0 \tag{18}$$

Considering the judgment criteria of temperature strength, "$l > 0.2$" samples are filtered and written into the data.txt file, which would be utilized for merging and selecting thermoclines.

Based on these factors combined with the high-resolution data, the existence of a thermocline in the area can be accurately judged. Then, the position of the thermocline's lateral boundary can be accurately determined according to the determination of the critical position of thermocline and non-thermocline area.

4.4 Trend prediction for the lateral boundary of thermocline
As the effectiveness of the SVR method in ocean data prediction is verified by the analysis above, the high-resolution data is obtained after refinement, and a method for determining the boundary of the thermocline is provided. Combining the existing BOA Argo data from 2004 to 2015, the 4-year trend prediction for the lateral boundary of the thermocline is conducted through the SVR method from 2016 to 2019. The temperature distribution of the real data, high-resolution, and the prediction results are compared in Fig. 5. In addition, the
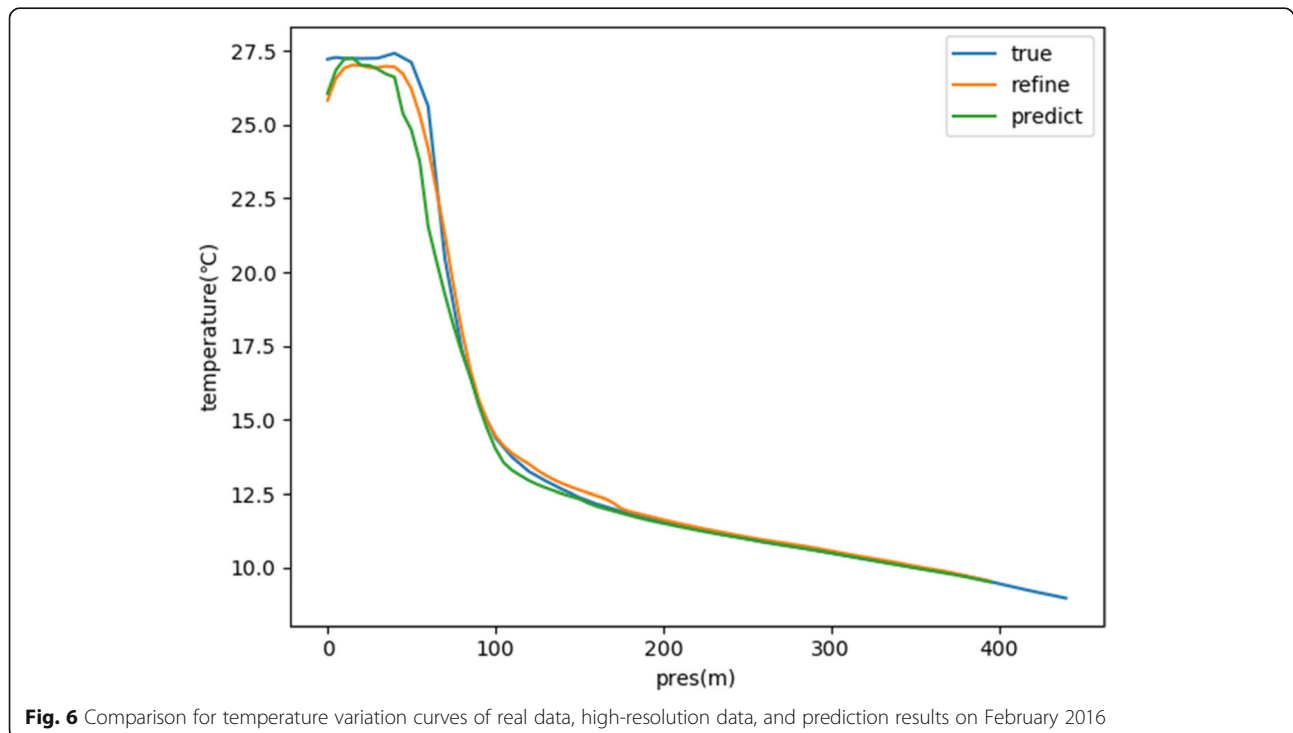
Fig. 6 Comparison for temperature variation curves of real data, high-resolution data, and prediction results on February 2016

temperature variation curves of the real data, refined data, and prediction results at a chosen area (57.5°E, 16.5°S) are compared as well in Fig. 6.

It can be known from Fig. 5 that the temperature distribution predicted by the SVR method is very similar to that under the high-resolution data, which is consistent with the overall temperature distribution trend of the real data. The temperature gradient boundary and distribution can be accurately obtained from the prediction results, which show that the use of SVR method is effective for forecasting the temperature boundary accurately in the horizontal direction. In Fig. 6, the trend of temperature changing curves for the real data, high-resolution data, and prediction results are basically the same, and the predictive temperature value is slightly lower than the true value merely in the condition of the shallow seawater. Therefore, it is reckoned that the temperature prediction results with high precision can

be obtained through the SVR method in the depth direction. By comparing the temperature distribution in the horizontal direction and the temperature variation curve in the depth direction, it can be seen that the SVR method can be used to predict the ocean temperature variation, and then predict the variation trend of the three-dimensional thermocline's lateral boundary.

Based on the contrastive analysis of the temperature distribution in 2016, the paper predicts the temperature distribution in this area from 2017 to 2019, and still takes the February with the obvious change of thermocline as an example. The prediction results are shown in Fig. 7.

Figure 7 reveals the variation trends of the lateral boundary of the thermocline have no great difference from 2016 to 2019, and the temperature near the equator is obviously higher than that away from the equator as a whole. The highest temperature appeared at approximately 10°–15° South and longitude 70°–80° East,

Fig. 7 Prediction of changes in the thermocline boundary in 2016 (**a**), 2017 (**b**), 2018 (**c**) and 2019 (**d**)
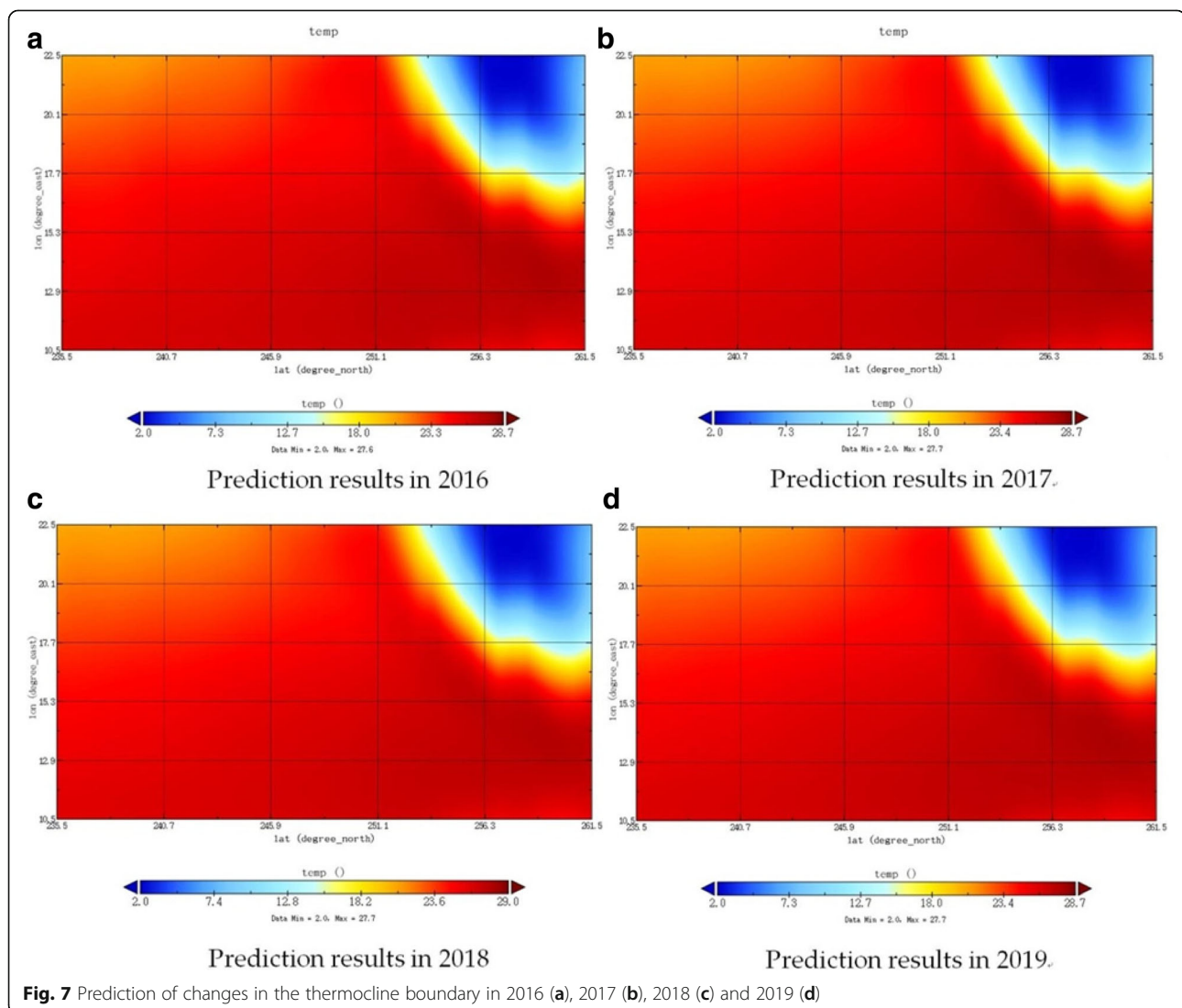
and the lowest temperature appeared at approximately 20°–22.5° South and longitude 75°–80° East. In this sea area, the maximum temperature reaches 29 °C in 2018, and the maximum temperature of 2019 is 28.7 °C, which is basically the same as that of 2016 and 2017. From 2016 to 2019, the range of high-temperature areas shows a decreasing trend, and the temperature gradient of the upper right part of the figure also decreases. Accordingly, it is obvious from the prediction results that the temperature change varies with the year, which can provide a reference for further research on the factors affecting the temperature change.

5 Conclusions

In this paper, the SVR method in machine learning is applied to predict the variation trend of three-dimensional thermocline's lateral boundary in the study area (10°–25° S and 55°–80° E). In this condition, the paper first utilizes the original temperature and depth data of the ocean to make a prediction of 2016 with SVR method, and compares them with the real data of 2016 in order to verify the feasibility of the SVR method. Based on the SVR method, the temperature and depth data of the sea area are then processed with high-resolution data (horizontal resolution of 0.01° and vertical resolution of 5 m). The changes in the high-resolution temperature distribution of refinement results are easier for observation and analysis. Finally, the "information entropy method" in machine learning is combined with the traditional judgment method to determine the lateral boundary of thermocline in this paper. Meanwhile, the SVR model is adopted to analyze the variation trend of three-dimensional thermocline's lateral boundary from 2017 to 2019. The results show that the use of SVR method can realize the variation trend prediction for three-dimensional thermocline.

With regard to ocean data refinement in the future, the SVR method are adopted to refine the temperature, salinity, and depth data at a higher resolution, and the amount of refined data will grow exponentially. In the research of thermocline judgment, we study the three-dimensional boundary of thermocline so as to judge the three-dimensional "temperature jump body" more accurately and plan to propose the concept of three-dimensional "temperature jump body." The study of three-dimensional "temperature jump body" will provide a more precise thermocline location for marine fishing ground distribution and acoustic communication research. In terms of the thermocline prediction, we will further consider the external environmental factors, climate change, anthropogenic effects, and other conditions simultaneously, to obtain more accurate prediction results for thermocline.

Availability of data and materials
Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Authors' contributions
HQ and ZD conceived and designed the experiments; YJ and ZD performed the experiments; YJ and CW analyzed the data; HQ and CW wrote the paper. All authors read and approved the final manuscript.

Competing interest
The authors declare that they have no competing interest.

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

Author details
[1]Science and Technology on Underwater Vehicle Technology, Harbin Engineering University, Harbin 150001, China. [2]College of Computer Science and Technology, Jilin University, Changchun 130012, China. [3]Department of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310037, China.

References
1. D. Yuan, Z. Zhang, P.C. Chu, et al., Geostrophic circulation in the tropical North Pacific Ocean based on Argo profiles. J. Phys. Oceanogr. **44**(2), 558–575 (2014)
2. X.F. Wu, Q.L. Zhang, Z.H. Liu, Annual and interannual variations of the Western Pacific Warm Pool volume and sources of warm water revealed by Argo data. Sci. China. (Series D). **57**(9), 2269–2280 (2014)
3. L. Cheng, J. Zhu, Uncertainties of the ocean heat content estimation induced by insufficient vertical resolution of Historical Ocean subsurface observations. J. Atmos. Ocean. Technol. **31**(6), 1383–1396 (2014)
4. C.H. Sun, Z.H. Liu, M.R. Tong, et al., The application of Argo data to water masses analysis in the Northwest Pacific Ocean. J. Mar. Sci., **10**(2), 1–13 (2008)
5. P.K. Bhaskaran, R.R. Kumar, R. Barman, et al., A new approach for deriving temperature and salinity fields in the Indian Ocean using artificial neural networks. J. Mar. Sci. Technol. **15**(2), 160–175 (2010)
6. B. Gu, V.S. Sheng, A robust regularization path algorithm for v-support vector classification. IEEE Trans. Neural. Netw. Learn. Syst. **28**(5), 1241–1248 (2017)
7. B. Gu, V.S. Sheng, K.Y. Tay, et al., Incremental support vector learning for ordinal regression. IEEE Trans. Neural. Netw. Learn. Syst. **26**(7), 1403–1416 (2015)
8. Argo Science Team.On the design and implementation of Argo: an initial plan for a global array of profiling floats. ICPO Report No.21. GODAE International Project office, Bureau of Meteorology, 1998
9. X. Jianping, *A exploration of global ocean argo observing [M]* (Ocean Press, Beijing, 2002)
10. S. Dong, S. Janet, T. Gille Sarah, et al., Southern Ocean mixed layer depth from ARGO float profiles. J. Geophys. Res. Oceans, vol. 113, C06013 (2008). https://doi.org/10.1029/2006JC004051
11. R.E. Hadfield, N.C. Wells, S.A. Josey, et al., On the accuracy of North Atlantic temperature and heat storage fields from Argo. J. Geophys. Res. Oceans. **112**(C1), C01009-[17pp] (2007)
12. S. Guinehut, P.Y.L. Traon, G. Larnicol, et al., Combining Argo and remote-sensing data to estimate the ocean three-dimensional temperature fields—a first approach based on simulated observations. J. Mar. Syst. **46**(1–4), 85–98 (2004)

13. Y.D. Resnyanskii, M.D. Tsyrulnikov, B.S. Strukov, et al., Statistical structure of spatial variability of the ocean thermohaline fields from Argo profiling data, 2005–2007[J]. Oceanology. **50**(2), 149–165 (2010)

14. G. Maze, H. Mercier, R. Fablet, et al., Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the North Atlantic Ocean. Prog. Oceanogr. **151**, 275–292 (2017)

15. T.U. Bhaskar, M. Ravichandran, R. Devender, An operational objective analysis system at INCOIS for generation of Argo value added products. Indian National Center for Ocean Information Services (INCOIS) (2007). http://www.incois.gov.in/documents/TechnicalReports/INCOIS-MOG-ARGO-TR-04-2007.pdf

16. H. Shigeki, O. Tsuyoshi, N. Tomoaki, A monthly mean dataset of global oceanic temperature and salinity derived from Argo float observations. Jamstec. Rep. Res. Dev. **8**, 47–59 (2008)

17. D. Roemmich, J. Gilson, The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo program. Prog. Oceanogr. **82**(2), 81–100 (2009)

18. F. Gaillard, E. Autret, V. Thierry, et al., Quality control of large Argo datasets. J. Atmos. Ocean. Technol. **26**(2), 337–351 (2010)

19. P. Joseph, A History of Thermocline Theory. Physical Oceanography. Springer New York, pp. 139–152 (2006)

20. X.H. Zhang, X.D. Zhang, L.I. Yan, Characteristic analysis and calculation of the ocean thermocline. Marine Forecasts, **28**(5), 69–76 (2011)

21. D. Laurent, K. Holland, et al., Deep diving behavior observed in yellowfin tuna, (*Thunnus albacares*). **19**(1), 85–88 (2006)

22. VLIZ IMIS, Romena N A. Factors affecting distribution of adult yellowfin tuna (Thunnus albacares) and its reproductive ecology in the Indian Ocean based on Japanese tuna longline fisheries and survey information. Minerva. Pediatr., 2000, 29(34):2027–2030

23. Song L M, Zhang Y U, LIU XIONG X U, et al. Environmental preferences of longlining for yellowfin tuna (Thunnus albacares) in the tropical high seas of the Indian Ocean. Fish. Oceanogr., 2008, 17(4):239–253

24. Z. Xu, Y.G. Zhang, Simulation for acoustic channel influenced by shallow water thermocline. J. Syst. Simul. **24**(10), 2167–2166 (2012)

25. B. Jiang, X.R. Wu, J. Ding, Comparison of the calculation methods of the thermocline depth of the South China Sea. Mar. Sci. Bull. **35**(1), 64–73 (2016)

26. Y. Jiang, Y. Gou, T. Zhang, et al., A machine learning approach to Argo data analysis in a thermocline. Sensors. **17**(10), 2225 (2017)

27. S. Yang, Y. Zhang, H. Zhang, et al., The relationship between the temporal-spatial distribution of fishing ground of yellowfin tuna (Thunnus albacares) and themocline characteristics in the tropic Indian Ocean. Acta. Ecol. Sin. **32**(3), 671–679 (2012)

28. M.C. Gregg, Diapycnal mixing in the thermocline: a review. J. Geophys. Res. Oceans. **92**(C5), 5249–5286 (1987)