China Ocean Eng., 2019, Vol. 33, No. 2, P. 172–184
DOI: 10.1007/s13344-019-0017-5, ISSN 0890-5487
http://www.chinaoceanengin.cn/ E-mail: coe@nhri.cn

A Marine Environment Early Warning Algorithm Based on Marine Data Sampled by Multiple Underwater Gliders

XU Zhen-zhen[a], LI Lu[a], YU Jian-cheng[b, *], XU Xiu-juan[a], JIA Ming-fei[a]

[a]School of Software Technology, Dalian University of Technology, Dalian 116620, China

[b]The State Key laboratory of Robotics, Shenyang Institute of Automation, CAS, Shenyang 110016, China

Received September 10, 2018; revised November 21, 2018; accepted December 6, 2018

Abstract

This study analyzes and summarizes seven main characteristics of the marine data sampled by multiple underwater gliders. These characteristics such as the big data volume and data sparseness make it extremely difficult to do some meaningful applications like early warning of marine environment. In order to make full use of the sea trial data, this paper gives the definition of two types of marine data cube which can integrate the big marine data sampled by multiple underwater gliders along saw-tooth paths, and proposes a data fitting algorithm based on time extraction and space compression (DFTS) to construct the temperature and conductivity data cubes. This research also presents an early warning algorithm based on data cube (EWDC) to realize the early warning of a new sampled data file. Experiments results show that the proposed methods are reasonable and effective. Our work is the first study to do some realistic applications on the data sampled by multiple underwater vehicles, and it provides a research framework for processing and analyzing the big marine data oriented to the applications of underwater gliders.

Key words: big marine data, early warning, marine environment, underwater gliders

Citation: Xu, Z. Z., Li, L., Yu, J. C., Xu, X. J., Jia, M. F., 2019. A marine environment early warning algorithm based on marine data sampled by multiple underwater gliders. China Ocean Eng., 33(2): 172–184, doi: 10.1007/s13344-019-0017-5

1 Introduction

Ocean is a treasure to human society since it is rich in various resources like marine biological resources (e.g. fish), marine mineral resources (e.g. petroleum) and marine energy (e.g. tidal power). Marine environmental monitoring is an important means to supervise and manage the marine environment, and it is the foundation of all the work related to the marine environment. The early warning about abnormal data in the marine environment is of great significance to marine environmental monitoring.

Conventional marine environmental monitoring measures include buoy, shore station and ship. However, none of the measures mentioned above can obtain continuous and high-resolution ocean monitoring data in a long period of time. In recent years, underwater glider, which is a new type of ocean environment observation vehicle, has received great attention of researchers.

Underwater glider can take a variety of sensors and move along a saw-tooth trajectory in a wide range with a long continuous period of time. It can conduct marine monitoring on ocean physical and chemical parameters (e.g. temperature, salinity and acoustic characteristics) for marine science research. In addition, the underwater glider also has such merits as low cost of manufacturing, strong endurance ability, independent control and so on (Zhang et al., 2011; Yu et al., 2013). Thus, underwater gliders have broad application prospects in the field of marine environmental monitoring.

It is impracticable for a single glider to get sample data from different locations at the same time, while multiple underwater gliders can avoid this limitation (Zhang et al., 2015). Underwater gliders dive into ocean, and then swim up to the ocean surface, which can provide us with both the vertical structure of ocean and continuous marine data. Underwater gliders have been successfully used to investigate characteristics and conditions of ocean in several regions of the world, such as the Northern South China Sea (Qiu et al., 2015), the Monterey Bay (Fiorelli et al., 2006), the Balearic Sea (Bouffard et al., 2010), and the Ionian Sea (Dobricic et al., 2010).

In this paper, the marine data are sampled in the South China Sea from June 2014 to June 2015 by several under-

Foundation item: This work is financially supported by the National Natural Science Foundation of China (Grant Nos. U1709202 and No. 61502069), the Foundation of State Key Laboratory of Robotics (Grant No. 2015-o03) and the Fundamental Research Funds for the Central Universities (Grant Nos. DUT18JC39 and DUT17JC45).

*Corresponding author. E-mail: yjc@sia.cn water gliders designed by Shenyang Institute of Automation, Chinese Academy Sciences. The observational data has a big data volume since the environmental parameters are recorded every six seconds. Meanwhile, to the vast ocean, the sampled data are highly sparse because gliders collect data along a saw-tooth trajectory. In addition, the sampled marine data have characteristics like multiple sources, multiple parameters and inconsistency at the time and space. Because of these characteristics, it is difficult to do the theoretical and practical research on the data sampled by multiple underwater gliders. It is significant to explore a marine environment early warning algorithm by making full use of these sea trial data. It can help the researchers and policymakers to analyze the data and find out the useful information hidden in these data.

Vasilijević et al. (2017) presented a cooperative robotic system for environmental monitoring consisting of an autonomous underwater vehicle (AUV) and an autonomous unmanned surface vehicle (USV), and a novel human-on-the-loop (HOTL) approach is applied on the system for environmental monitoring. Båmstedt and Brugel (2017) proposed a cost-precision model for marine environmental monitoring based on the time-integrated averages. The environmental data sampled by many sampling stations in the northern Bothnian Sea were used in this cost-precision spatio-temporal allocation model. A Marine Information System, acting as an integrated and inter-operable monitoring tool is proposed and discussed by Pieri et al. (2018). However, the data from AUV reports are acquired only on specific occasions in this system. All the above monitoring systems are not based on multiple gliders. Thus, we should set up a new model to monitor the marine environment oriented to multiple gliders.

Forecasting and early warning methods have a wide range of applications and abundant achievements in many fields. Zheng et al. (2012) studied safety evaluation and early warning rating of the hot and humid environments, and proposed a fuzzy analytic hierarchy process (AHP) method to evaluate the work safety in hot and humid environments. Sun and Lee proposed a red tide prediction method that uses the fuzzy reasoning and ensemble method to forecast the density of red tide algae and red tide blooms (Park and Lee, 2014). Park et al. (2015) used artificial neural network and support vector machine to predict the concentration of *chlorophyll-a* for the early warning in freshwater and estuarine reservoirs. Fang et al. (2015) presented an integrated approach to snowmelt floods early-warning based on geoinformatics (i.e. remote sensing, geographical information systems and global positioning systems), Internet of Things (IoT) and cloud services. Zollo et al. (2010) proposed an integrated regional/on-site early warning method, which can be used in the very first seconds after a moderate-to-large earthquake to map the most likely damaged zones. Jiang et al. (2016) built a novel framework based on a principal component analysis and an improved continuous hidden Markov model for forecasting and early warning of microcystins.

All the research work mentioned above are oriented to certain application fields. However, there is no relevant research on the early warning of the ocean environment based on sampled data of multiple underwater gliders. The contributions of this paper include:

(1) Analyze the characteristics of marine data sampled by multiple underwater gliders in the South China Sea.

(2) Define the concept of marine data cube to describe the marine data, and propose a data fitting algorithm DFTS to construct the marine data cube based on large but sparse sampled data.

(3) Propose an early warning algorithm based on the marine data cube to realize the early warning of new sampled data.

To the best of our knowledge, this paper is the first attempt to conduct an early waring application based on the marine data sampled by multiple underwater gliders. Section 2 analyzes seven main characteristics of sampled marine data and introduces the preprocessing procedure of these data. Section 3 defines the concept of marine data cube and proposes a data fitting algorithm based on the time extraction and space compression (DFTS) to construct the marine data cube. In Section 4, an early warning algorithm based on data cube (EWDC) is described in detail. Plenty of experiments are carried out to determine the appropriate parameters of EWDC in Section 5 and the last section is the conclusion.

2 Data analysis and preprocessing

There are six data packages observed by underwater gliders from June 2014 to June 2015 in the South China Sea. The basic information of all data packages is shown in Table 1 including the experimental period, longitude and

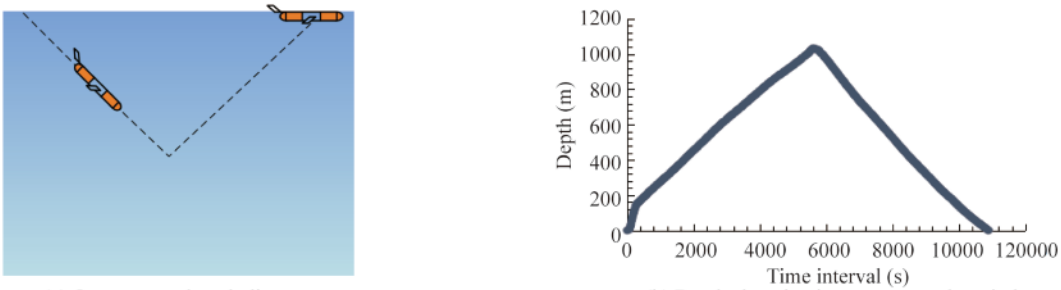
Table 1 Basic information of all data packages

Data package ID	Experimental period	Longitude range (GPS data format)	Latitude range (GPS data format)	Number of CTD data files	Records of CTD data files
1	Jun. 6, 2014 –Jun. 26, 2014	11456.903–11852.914	1847.211–2124.178	126	246514
2	Sep. 10, 2014 –Oct. 15, 2014	11628.464–11857.39	1941.643–2137.226	227	385760
3	Nov. 13, 2014 –Nov. 17, 2014	11356.853–11857.289	1803.938–2136.889	38	61560
4	Apr. 18, 2015 –May. 7, 2015	11033.03–11131.14	1656.744–1754.551	138	467854
5	Apr. 28, 2015 –May. 21, 2015	11628.464–11857.390	1941.643–2137.226	144	259919
6	Apr. 28, 2015 –Jun. 1, 2015	11356.853–11857.289	1803.938–2136.889	205	299112

latitude range, the number of CTD (Conductivity–Temperature- Depth) data files and the number of data records in corresponding CTD data files.

A CTD data file records the data collected by an underwater glider at one saw-tooth period, which contains several marine monitoring parameters including the time interval, the depth, temperature, conductivity and the task time. Every six seconds, the CTD data file adds a record which contains all these parameters (Xu et al., 2016). Fig. 1 gives an example of the path and depth changing in one CTD data file, and we can see the data is from one saw-tooth period.

There is also a task information file in each package which records the task information of each CTD data file in this package. These parameters include the task name, the date, time, type and the operator name, as well as some gliders' parameters like the longitude and latitude of the starting and ending positions, desired and real maximum depth, desired heading and pitch angle, leakage voltage, periodicity range (i.e. the spherical distance between the start and end positions), and so on.



(a) One saw-tooth path diagram

(b) Depth changing in one saw-tooth period

Fig. 1. An example of a CTD data file.

We summarize seven main characteristics of the sampled data:

(1) Big data volume: Each package contains tens of or even hundreds of CTD data files and one experimental task information file. The total files number is up to 884 and the number of data records is up to 1720719.

(2) Data sparseness: Underwater gliders move along a saw-tooth trajectory and collect the marine observational data. Compared with broad ocean, the sampled marine data are still extremely limited and sparse.

(3) Multiple sources: The marine data are collected by several underwater gliders in different regions and distinctive time, thus making the monitoring more comprehensive than that using buoy and shore station.

(4) Multiple parameters: As is mentioned above, two types of files are contained in a packet and each of them consists of multiple parameters.

(5) Time inconsistency: The sea trials are conducted during a long period of time, from June 2014 to June 2015. Each data package belongs to a distinct experimental period. In a certain package, the experimental time of tasks differs from each other.

(6) Position inconsistency: The route of each task is different from others. Although the last three packages seem to overlap in time, indeed, they are conducted by three different gliders in different places, respectively. There are no such two tasks conducted in the same area simultaneously. Although some points from different routes are overlapped in the longitude and latitude, they are not overlapped in the depth at the same time. Thus, the real positions of each glider in these routes are quite different.

(7) Inconsistency or loss of data: The Autonomous Underwater Gliders work automatically, so it is inevitable that the faults on the sensor may result in the data loss and inconsistency. For example, the task conducted on April 29, 2015 by Glider SIA-G1000J003 records an impossible maximum depth as deep as 24021 m.

Due to the characteristics of sampled marine data introduced above, it is difficult to effectively use existing observational data to realize the early warning of the marine environment. First of all, the data rectification should be conducted to obtain valid data. The GPS data format is not easy to be used in the following data fitting algorithm. So, the GPS data format needs to be transformed to the format with the unit of degree. In addition, marine data collected by the underwater gliders may contain null values and wrong values. The values beyond normal ranges should be regarded as wrong values and be replaced by the modified ones, which are the average value of the previous and next data items. After the outlier filtering, the data are valid and can be used to solve the early warning problem.

3 Data cube construction

3.1 Definition of the data cube

In general, the definition of a data cube is a multidimensional data set in the data mining (van der Aalst, 2013). In this paper, we define two types of marine data cube which are called TDC (temperature data cube) and CDC (conductivity data cube). Let M denotes the month, and a TDC is defined as a 4-tuple form which is composed of the longitude, latitude, depth and temperature with a given M. The formulation is shown as follows:

$$TDC_M =< LNG, LAT, DEP, TEMP >, \quad M_1 \leqslant M \leqslant M_k, \quad (1)$$

where *LNG*, *LAT*, *DEP* and *TEMP* denote the longitude, latitude, depth and temperature, respectively. Seawater is divided into eighteen layers from 0 to 1000 meters according to the standard observation level, i.e., let *L* presents the set {0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 1000}, $DEP \in L$. Similarly, a CDC is defined as a 4-tuple form including the longitude, latitude, depth and conductivity with a given month value, which is shown in Eq. (2). The value range of *DEP* is the same with the definition in the TDC. In the two formulas, *k* denotes the total number of the months on which the sea trials have carried out.

$$CDC_M =< LNG, LAT, DEP, COND >, \quad M_1 \leqslant M \leqslant M_k. \quad (2)$$

3.2 Data fitting algorithm DFTS

In order to establish the data cube to display the big marine data obtained by the underwater gliders, we propose a data fitting algorithm based on the time extraction and space compression (DFTS). We focus on the data fitting of the temperature and conductivity data. The main idea of the DFTS is based on the following two aspects.

(1) Take the month as the particle size of time in the data fitting, i.e., we conduct the data fitting for each month separately.

(2) Take 18 depth layers as the particle size of space in the data fitting. That means we filter the data around each layer and then fit it.

The data fitting algorithm DFTS includes five steps. Taking the data sampled in April 2015 as an example, the detailed procedure of the DFTS is described as follows:

(1) Extracting from the task information files by month. We extract the data including the longitude and latitude of the start point (i.e., In-Longitude and In-Latitude), longitude and latitude of the end points (i.e., Out-Longitude and Out-Latitude), and the CTD file names from those records whose date is April 2015 in all task information files. Then input these data into a new file named TIF-201504. Fig. 2 gives the longitude and latitude values of extracted data in April 2015 which is actually composed of three gliders' paths.

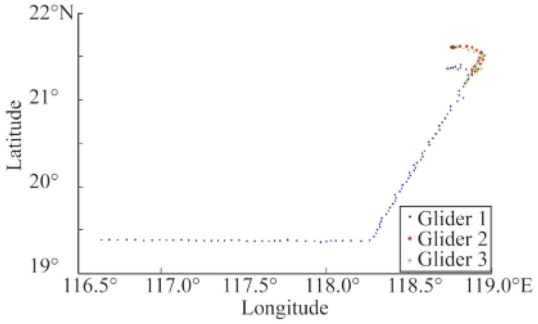


Fig. 2. Longitude and latitude values of extracted data in April 2015.

(2) Horizontal space compression. Calculate the average value of In-Longitude and Out-Longitude, and the average value of In-Latitude and Out-Latitude. The average values are noted as AVE-LNG and AVE-LAT and added into the file TIF-201504. This step compresses one saw-tooth range into one point because the distance of the start and end points is relatively small and the temperature and conductivity have no obvious change in such a limited range. Table 2 gives partial data of TIF-201504 after the horizontal space compression.

(3) Extracting from the CTD data files by month. Find the CTD data files whose names are on the list of the file TIF-201504. Then, add AVE-LNG and AVE-LAT into each record of corresponding CTD data file. After that, integrate all the CTD data files in the same month into one single file named CTD-201504.

(4) Vertical space compression. Divide the data into eighteen layers according to the depth. Define a range of (*DEP*–1, *DEP*+1) for each layer except the layer of 0 m. Especially, we define the depth range from 0 m to 1.5 m as the layer of 0 m. After following the above steps, examples of the data in the layer of 0 m and 10 m are shown in Table 3 and Table 4.

(5) Fitting with the polynomial linear model. We fit the temperature and conductivity data of each layer with the polynomial linear model. The experimental results will be introduced in Section 3.3.

According to the algorithm description mentioned above, the pseudo-code of the DFTS is shown in Algorithm 1.

3.3 Data cube construction

Fig. 3–Fig. 5 show the data fitting results of the temper-

Table 2 Partial data of TIF-201504 after the horizontal space compression

In-LNG (°E)	Out-LNG (°E)	In-LAT (°N)	Out-LAT (°N)	AVE-LNG (°E)	AVE-LAT (°N)	File name
118.73	118.73	21.36	21.36	118.7326	21.3631	150418_2_1
118.73	118.75	21.36	21.37	118.7429	21.3637	150418_3_1
118.75	118.77	21.36	21.37	118.7577	21.3686	150418_4_1
118.76	118.78	21.37	21.38	118.7718	21.3776	150418_5_1
118.78	118.82	21.38	21.40	118.7976	21.3938	150418_6_1
118.82	118.80	21.40	21.35	118.8063	21.3761	150418_7_1
118.80	118.81	21.35	21.37	118.8043	21.3605	150418_8_1
118.81	118.85	21.37	21.35	118.8290	21.3618	150419_1_1
118.85	118.88	21.35	21.33	118.8639	21.3424	150419_2_1

Table 3 Partial data in the layer of 0 m

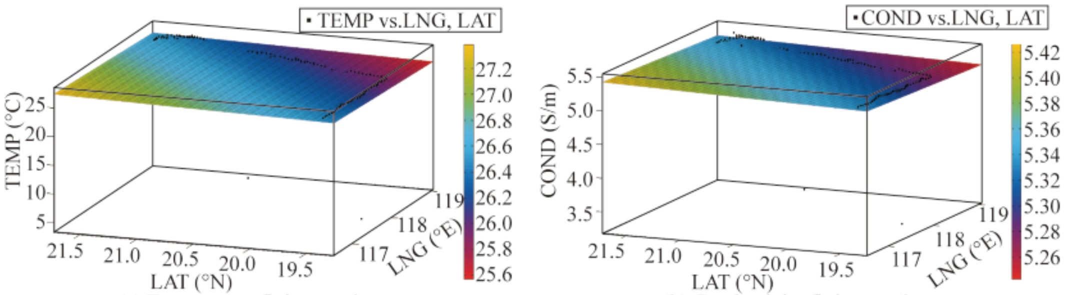
COND (S/m)	TEMP (°C)	Depth (m)	AVE-LNG (°E)	AVE-LAT (°N)
5.37770	26.174	1.1	118.7325583	21.363067
5.37668	26.164	1.2	118.7325583	21.363067
5.37640	26.168	1.3	118.7325583	21.363067
5.37561	26.160	1.3	118.7325583	21.363067
5.37475	26.150	1.4	118.7325583	21.363067
5.36030	26.036	1.5	118.7325583	21.363067
5.37517	26.186	1.1	118.7429333	21.363700
5.37316	26.170	1.2	118.7429333	21.363700
5.36963	26.128	1.3	118.7429333	21.363700

Table 4 Partial data in the layer of 10 m

COND (S/m)	TEMP (°C)	Depth (m)	AVE-LNG (°E)	AVE-LAT (°N)
5.34538	25.894	9.7	118.73256	21.36307
5.33346	25.798	10.1	118.73256	21.36307
5.29946	25.682	9.6	118.77184	21.37764
5.28493	25.532	9.9	118.77184	21.37764
5.29848	25.648	10.4	118.79755	21.39380
5.23204	25.172	10.1	118.79755	21.39380
5.23572	25.190	9.8	118.80628	21.37613
5.29873	25.638	10.0	118.80628	21.37613
5.29177	25.720	9.6	118.80425	21.36048

Algorithm 1. Data fitting algorithm based on time extraction and space compression (DFTS)

```
Input: All preprocessed files including CTD data files and task information files
Output: temperature data cube $TDC_M$ and conductivity data cube $CDC_M$
DFTS(){
For M = $M_1$ : $M_k$
    Extract the data from task information files in month M into file TIF-M
    For $i$ = 1 : Length (TIF-M)
        AVE-LNG = (In-Longitude + Out-Longitude) / 2
        AVE-LAT = (In-Latitude + Out-Latitude) / 2
        Add AVE-LNG and AVE-LAT into each record of corresponding CTD data file
    End for
    Integrate all the CTD data files in month M into File CTD-M
    L = {0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 1000}
    For DEP = L[1] : L[18] //Divide the records of File CTD-M into eighteen layers according to the depth
        If DEP = L[1]
            Compress Depth Range = [0, 1.5]
        Else
            Compress Depth Range = [DEP - 1, DEP + 1]
        End if
        Fit (AVE-LNG, AVE-LAT, TEMP) to get temperature surface of DEP meters
        Fit (AVE-LNG, AVE-LAT, COND) to get conductivity surface of DEP meters
    End for
    Combine 18 surfaces together to get $TDC_M$ and $CDC_M$
End for
}
```



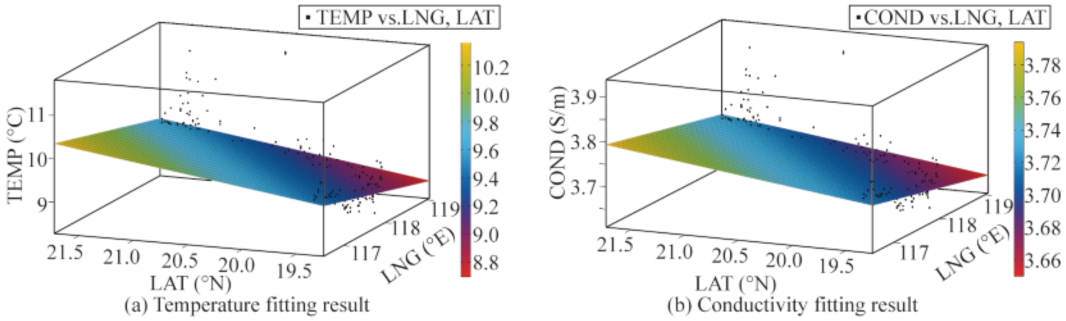
(a) Temperature fitting result

(b) Conductivity fitting result

Fig. 3. Results of the data fitting in the layer of 10 m.

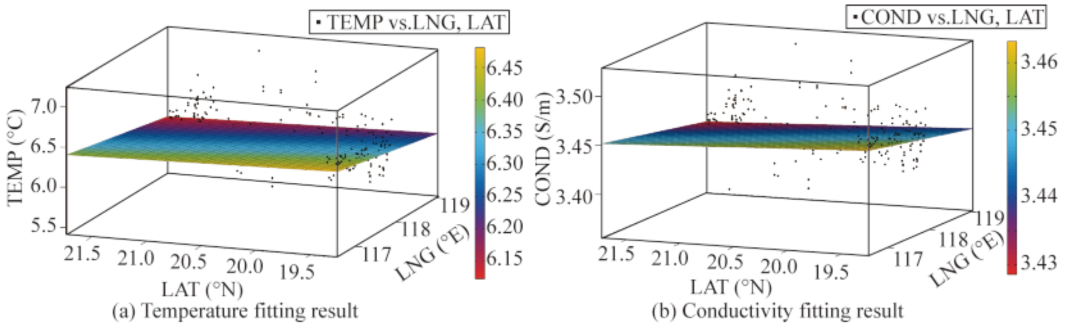
ature and conductivity in the layers of 10 m, 400 m, and 700 m, respectively. In these figures, the dark points stand for the data of the temperature and conductivity in the layer range, and the colorful surface represents the fitting surface.

Since the depth of the marine data is divided into eighteen layers, there are eighteen fitting surfaces in total. According to the definition of the marine data cube, all the fitting surfaces can be combined together to display the big marine data sampled by multiple underwater gliders in the same month. Figs. 6a and 6b show the temperature data cube and conductivity data cube in April 2015, respectively. We can see from Fig. 6 that the proposed data fitting al-



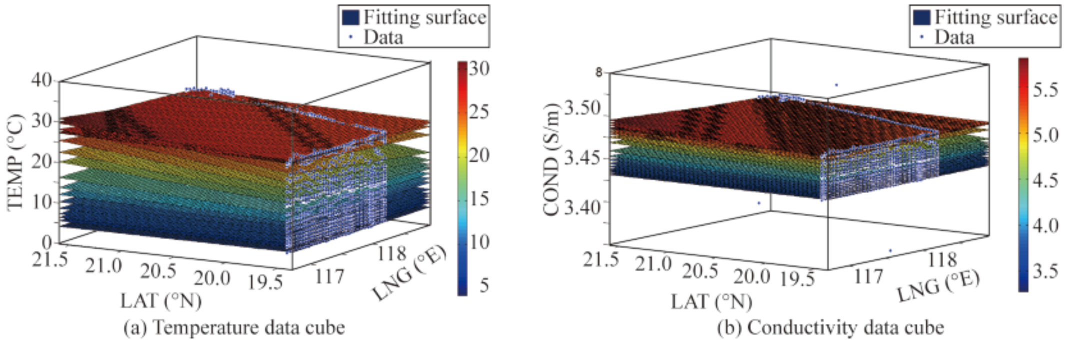
(a) Temperature fitting result (b) Conductivity fitting result

Fig. 4. Results of the data fitting in the layer of 400 m.



(a) Temperature fitting result (b) Conductivity fitting result

Fig. 5. Results of the data fitting in the layer of 700 m.



(a) Temperature data cube (b) Conductivity data cube

Fig. 6. Marine data cube.

gorithm DFTS is feasible to construct the data cube. However, in the depth of 0 to 75 m, some fitting surfaces may be overlapping because of the instability of the data. And the fitting effect varies according to the layer depth. In the next step, we are going to study the fitting effect of each layer by the quantitative analysis.

This study use the root mean square error (RMSE) (Chai and Draxler, 2014) to compare the fitting effect of the temperature and conductivity on each fitting surface. It is known that smaller RMSE means better data fitting results. The formula is expressed as:

$$RMSE = \sqrt{\frac{\sum d_i^2}{m}}, \quad i = 1, 2, \cdots, m, \tag{3}$$

where d_i is the deviation of the real value and fitting value on each fitting surface, and m is the number of real values.

The relationship between the depth and RMSE of the temperature in April 2015 is shown in Fig. 7a. During the shallow water of the layer of 0 m to the layer of 50 m, the RMSE decreases sharply from 2.16 at the layer of 0 m to 0.7262 at the layer of 20 m, and then goes back to 1.274 at the layer of 50 m. The reason is the data collected by underwater gliders in the shallow water is not stable. From the layer of 50 m, the RMSE shows an overall downward trend. However, the RMSE has a slight increase to 1.042 at the layer of 125 m because the change of the temperature is relatively large in the ocean thermocline (Dong et al., 2015). The RMSE continuously decreases when the depth is more than 125 m, and reaches 0.139 when the depth is 1000 m. On the whole, the RMSE of temperature is smaller than 1.

Fig. 7b shows the relationship of depth and RMSE of conductivity in April 2015. The RMSE sharply goes down from 0.4723 at layer 0 m to 0.0683 at layer 20 m. Then, the RMSE increases until reaching the peak value of 0.1203 at

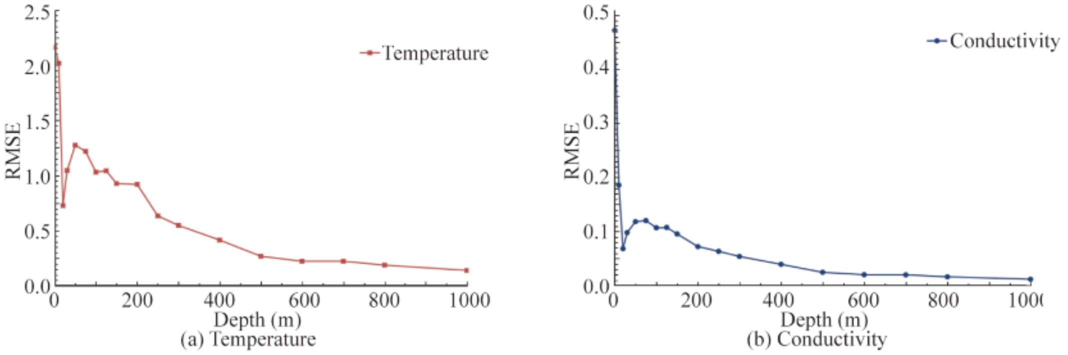


Fig. 7. Relationship between the depth and RMSE.

layer 75 m. When the depth is deeper than 75 m, the RMSE shows a downtrend on the whole, although the RMSE increases a little from 0.1071 at layer 100 m to 0.1077 at layer 125 m. When the depth exceeds 125 m the RMSE continuously decreases and reaches 0.0119 at the depth of 1000 m. In general, the RMSE of conductivity is smaller than 0.1.

Due to the different units and orders of the magnitude of the temperature and conductivity, we use Z-score (i.e., standard score) (Cheadle et al., 2003) to normalize the RMSE of the temperature and conductivity in order to compare the fitting effects. The formula of computing *Z*-score of the temperature is as follows:

$$Z_{\mathrm{T}} = \frac{\left(RMSE_{\mathrm{T}}^{p} - \mu_{\mathrm{T}}\right)}{\sigma_{\mathrm{T}}}, \qquad p = 1,2,\cdots,q, \tag{4}$$

where q represents the number of water layers, i.e. $q = 18$. $RMSE_{\mathrm{T}}^{p}$ denotes the temperature RMSE value in Layer p. μ_{T} is the mean value and σ_{T} is the standard deviation of q temperature RMSE values, which can be computed by Eq. (5) and Eq. (6), respectively.

$$\mu_{\mathrm{T}} = \frac{\sum_{p=1}^{q} RMSE_{\mathrm{T}}^{p}}{q}; \tag{5}$$

$$\sigma_{\mathrm{T}} = \sqrt{\frac{\sum_{p=1}^{q}\left(RMSE_{\mathrm{T}}^{p} - \mu_{\mathrm{T}}\right)^2}{q-1}}. \tag{6}$$

The formulas to compute the *Z*-score of the conductivity are similar with the Eqs. (4)–(6). After normalizing the RMSE of the temperature and conductivity, we draw the figure of the normalized RMSE values, which is shown in Fig. 8.

The distance between the normalized data and the *X* axis (i.e., $|Z_{\mathrm{T}}|$ and $|Z_{\mathrm{C}}|$) can be acted as the comparison criteria of the fitting effect. A smaller distance means better fitting effect. As can be seen from Fig. 8, in the layer of 0 m, 20 m and 200 m, $|Z_{\mathrm{T}}|$ values are 2.2498, 0.1803 and 0.1460 while $|Z_{\mathrm{C}}|$ values are 3.5916, 0.2461 and 0.2089, respectively. This means that in these three layers, the fitting effect of the temperature is better than that of the conductivity. But in other layers, $|Z_{\mathrm{C}}|$ values are all smaller than $|Z_{\mathrm{T}}|$ values. So, the fitting effect of the conductivity is better than that of the temperature in most cases.

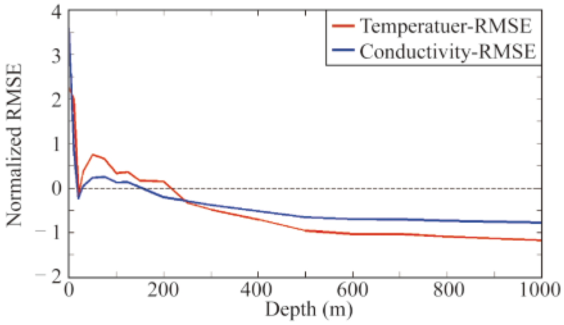


Fig. 8. Relationship between depth and normalized RMSE value.

4 Early warning method EWDC

4.1 Problem description

It is very important to detect the anomalies in monitoring marine environment. In this paper, the objective of early warning is finding out the abnormal temperature or conductivity values in a newly sampled CTD data file based on the historical data and output the specific coordinates (the longitude, latitude and depth) of the abnormal data. It is helpful for researchers and policymakers to make further studies and decisions after they obtain the early warning results.

4.2 Algorithm description

We have constructed the data cube of the marine information in the preceding section, which means we can obtain the temperature and conductivity information at any location of the 18 layers within the data cube. Then we can predict the temperature and conductivity data at any location in the cube range. Thus, when the newly sampled experimental data is given, we can detect whether there is abnormal information in it based on the data cubes. Thus, we present an early warning algorithm based on the data cube (EWDC).

With the changes of the seasons, temperature and conductivity will change accordingly. So it is reasonable to judge the correctness of sampled temperature and conductivity values by comparing the new test data with the historical data sampled at the same month. The main idea of the EWDC is matching the month of the data cube with the month of new CTD file, and exploring the abnormal data in the new test data. This algorithm has the following six steps.

Step 1. Preprocess the test values. A new CTD data file should be preprocessed according to the regulations described in Section 2. After that, calculate AVE-LNG (the average value of In-Longitude and Out-Longitude) and AVE-LAT (the average value of In-Latitude and Out-Latitude) and add the two parameters into the CTD data file. Thus, we can get information including AVE-LNG, AVE-LAT, depth, temperature, conductivity and month from the new CTD file.

Step 2. Match the month. Find out the existing data cubes including TDC and CDC whose month value is equal to the month of the new CTD file.

Step 3. Match the coordinates. Determine whether the AVE-LNG and AVE-LAT values of the new CTD data file are in the range of the data cubes found in Step 2. If yes, go to Step 4, output "early warning failed" prompt message otherwise.

Step 4. Extract the data cube values. Get the temperature and conductivity values from the data cube according to the coordinate value in the new CTD file. Since there are 18 layers, the corresponding 18 points' marine information (temperature and conductivity) can be obtained from the data cube.

Step 5. Compute the predictive values. We want to find out the relationship between the temperature and depth (or conductivity and depth), where the temperature or conductivity is the dependent variable, and the depth is the independent variable. Regression analysis can estimate the relationship between two variables. The most frequently used regression methods include linear regression, polynomial regression, ridge regression and lasso regression. Ridge regression and lasso regression are biased estimation regression methods specially used for multicollinearity data (high correlation between multiple independent variables). Here is only one independent variable. From the scatter figure, the relationship between the temperature (or conductivity) and depth is not a simple linear correlation. So we use the polynomial regression to fit the relationship between the depth and temperature (or depth and conductivity) from 18 points. The detailed algorithm for choosing the order of the regression equation is described in Section 4.3. Then all depth values in the new CTD file are brought into the equation to obtain the corresponding results of the temperature and conductivity data.

Step 6. Output the early warning results. Compare the predicted values with the test value in the new CTD data to get the difference. If the absolute value of the difference is larger than a threshold value, the early warning results with the detailed information including the longitude, latitude, depth and the difference with the normal values will be outputted. Otherwise, it is regarded as the normal data, and shows a "Normal" prompt message.

The pseudo-code of the EWDC is shown in Algorithm 2.

4.3 Determining the order of equation

According to Step 5 of the EWDC, in order to compute the predictive values, we should firstly get the multiple linear regression equation from the fitting curve with 18 points. After constructing a data cube, we can obtain the longitude and latitude ranges of this data cube. Given a certain coordinate within the longitude and latitude ranges, we can obtain corresponding 18 temperature and conductivity values represented by $T_1, T_2, \cdots, T_{18}$ and $C_1, C_2, \cdots, C_{18}$, respectively.

Polynomial regression is a special type of the multiple linear regression model, which is used widely in statistics (Montgomery et al., 2015). This paper uses this model to fit the 18 points' temperature and conductivity data. The matrix expression of the multiple linear regression model about the temperature is described as:

$$\boldsymbol{Y}_{\mathrm{T}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathrm{T}} + \boldsymbol{e}_{\mathrm{T}}, \tag{7}$$

where $\boldsymbol{Y}_{\mathrm{T}} = \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{pmatrix}_{n\times 1}$, $\boldsymbol{X} = \begin{pmatrix} 1 & L[1] & L[1]^2 & \cdots & L[1]^k \\ 1 & L[2] & L[2]^2 & \cdots & L[2]^k \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & L[n] & L[n]^2 & \cdots & L[n]^k \end{pmatrix}_{n\times(k+1)}$.

$\hat{\boldsymbol{\beta}}_{\mathrm{T}}$ is the estimated value of the regression coefficient, $\hat{\boldsymbol{\beta}}_{\mathrm{T}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}_{\mathrm{T}}$, and e_{T} is the residual.

The sample regression equation (i.e. the regression line of temperature) is expressed as:

$$\hat{\boldsymbol{Y}}_{\mathrm{T}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{\mathrm{T}}, \tag{8}$$

where $\hat{\boldsymbol{Y}}_{\mathrm{T}} = \begin{pmatrix} \hat{T}_1 \\ \hat{T}_2 \\ \vdots \\ \hat{T}_n \end{pmatrix}_{n\times 1}$, $\hat{T}_i$ represents the estimated value of T_i. In addition, $\sum_{i=1}^{n} {e_{\mathrm{T}}}^2 = \sum_{i=1}^{n} \left(T_i - \hat{T}_i\right)^2$.

Set $L = \{0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 1000\}$ and $n = 18$.

Let R represents the rank of a matrix, $\hat{\boldsymbol{\beta}}_{\mathrm{T}}$ is existed only when Eq. (9) is satisfied.

$$R(\boldsymbol{X}) \geqslant k+1. \tag{9}$$

The order of the polynomial regression model can be obtained by the following three ways.

(a) Adjusted R^2.

R^2 and adjusted R^2 of the regression analysis are used to detect the goodness of fit of a model. R^2 is known as the coefficient of determination, and its range is 0–1. If R^2 is close to 1, the goodness of fit of the model is better. Let R^2_{T} represents the coefficient of determination of the temperat-

Algorithm 2. Early warning algorithm based on data cube (EWDC)

```
Input: data cubes, a new CTD data file F sampled in the month $M_F$ and a corresponding task information file
Output: early warning results (LNG, LAT, DEP, $\Delta$T) and (LNG, LAT, DEP, $\Delta$C)
EWDC(){
For $i = 1$ : length (F)
  Preprocess the new CTD data file to get (TEMP, COND, Depth, AVE-LNG, AVE-LAT, $M_F$)
End for
For M = $M_1$ : $M_k$
  If M = $M_F$
    Find out $TDC_M$ and $CDC_M$
    break
  End if
End For
If (AVE-LNG, AVE-LAT) is not in the range the data cubes $TDC_M$ and $CDC_M$
  Output "early warning failed" prompt message
Else
  Get 18 temperature values from $TDC_M$ at coordinate (AVE-LNG, AVE-LAT)
  Get 18 conductivity values from $CDC_M$ at coordinate (AVE-LNG, AVE-LAT)
  Get the temperature equation from fitting the curve with 18 temperature values
  Get the conductivity equation from fitting the curve with 18 conductivity values
  For $i = 1$ : Length (F)
    Bring Depth(i) into the temperature equation to obtain the predictive temperature value $TEMP_p(i)$
    Bring Depth(i) into the conductivity equation to obtain the predictive conductivity value $COND_p(i)$
    $\Delta T = | TEMP(i) - TEMP_p(i)|$
    If $\Delta T > H_T(i)$
      Warning and Output (AVE-LNG, AVE-LAT, Depth(i), $\Delta$T)
    End if
    $\Delta C = | COND(i) - COND_p(i)|$
    If $\Delta C > H_C(i)$
      Warning and Output (AVE-LNG, AVE-LAT, Depth(i), $\Delta$C)
    End if
  End for
End if
}
```

ure. The detailed formulas are expressed as follows:

$$R_{\mathrm{T}}^2=\frac{ESS_{\mathrm{T}}}{TSS_{\mathrm{T}}}=1-\frac{RSS_{\mathrm{T}}}{TSS_{\mathrm{T}}}; \tag{10}$$

$$TSS_{\mathrm{T}}=\sum(T_i-\bar{T})^2; \tag{11}$$

$$ESS_{\mathrm{T}}=\sum(\hat{T}_i-\bar{T})^2; \tag{12}$$

$$RSS_{\mathrm{T}}=\sum(T_i-\hat{T}_i)^2=\sum {e_{\mathrm{T}i}}^2, \tag{13}$$

where $\bar{T}$ is the average value of $T_1, T_2, \ldots, T_n$, i.e. $\bar{T}=\sum_{i=1}^{n} T_i \Big/ n$.

In application, if the regression equation adds a variable, R^2 tends to increase. But in fact, the goodness of fit is not related to the increase of R^2 due to this reason. So R^2 is not a suitable criterion to compare the goodness of fit between the multiple linear regression models, and it should be adjusted. Adjusted R^2 is known as the adjusted coefficient of determination, which is very useful in evaluating and comparing the regression models. The function of adjusted R^2 is to exclude the influence of the number of variables on the goodness of fit and to prevent over-fitting.

Let $\bar{R}_{\mathrm{T}}^2$ represent the adjusted coefficient of determination of the temperature which is formulized in Eq. (14).

$$\bar{R}_{\mathrm{T}}^2=1-\frac{RSS_{\mathrm{T}}/(n-k-1)}{TSS_{\mathrm{T}}/(n-1)}. \tag{14}$$

Adjusted R_{T}^2 and R_{T}^2 have the following relationship:

$$\bar{R}_{\mathrm{T}}^2=1-(1-R_{\mathrm{T}}^2)\frac{n-1}{n-k-1}. \tag{15}$$

(b) Akaike Information Criterion (AIC)

Akaike Information Criterion (AIC) is a criterion for model selection, and it can balance the goodness of fit of the model and the complexity of the model. The AIC can be expressed as:

$$AIC_{\mathrm{T}}=\log\left(\frac{\sum e_{\mathrm{T}}^2}{n}\right)+\frac{2k}{n}. \tag{16}$$

Considering the residual, AIC severely punishes the additional order of independent variable. The preferred model is the one with the lowest AIC value when selecting from a set of models.

(c) Bayesian Information Criterion (BIC)

Bayesian Information Criterion (BIC) is proposed by Schward in 1978, and it has similar effects with AIC for model selection. Both AIC and BIC attempt to resolve overfitting through adding a penalty term for the number of parameters in the model, but the penalty term in BIC is larger than that in AIC. Given a set of candidate models for the data, the one with the minimum BIC value is preferred. The formula for BIC is as follows:

$$BIC_{\mathrm{T}} = \log\left(\frac{\sum e_{\mathrm{T}}^{2}}{n}\right) + \frac{k\log n}{n}. \tag{17}$$

5 Experiments

5.1 Experiments about the equation

In order to verify the effect of proposed EWDC algorithm, we need to do experiments based on the data cube constructed by real sea trial data. We choose the data cube in September 2014 as an example. First, construct the data cube in September 2014 based on the DFTS algorithm. Then, get the longitude and latitude ranges of this data cube which are 110.5790°E–111.3876°E and 17.1898°N–17.9201°N, respectively. 50 coordinates within the above longitude and latitude ranges are selected randomly to test the three criterions, and the 50 experiments all obtain optimal order of 4 in both the temperature and conductivity. Therefore, in this paper, we choose $\bar{R}_{\mathrm{T}}^{2}$ as the main criterion, and the other two criteria are used as verification methods.

We design a new CTD data file which acts as the test data, and the partial data of this file after preprocessing according to Step 1 of EWDC are shown in Table 5. Moreover, the test data pass Step 2 and Step 3 of the EWDC because the month information is September, and AVE-LNG (110.5963°E) and AVE-LAT (17.5372°N) are in the range of the data cube in September 2014.

Table 5 Partial data of preprocessed new CTD file

No.	Conductivity	Temperature (°C)	Depth (m)	AVE-LNG (°E)	AVE-LAT (°N)	Month
1	4.901	30.0240	0	110.5963	17.5372	9
2	5.6020	30.0220	0.7	110.5963	17.5372	9
3	5.6021	30.0220	0.7	110.5963	17.5372	9
...	...	...	...	...	...	...
590	3.2945	4.3880	1005.2	110.5963	17.5372	9
...	...	...	...	...	...	...
966	5.5800	29.9040	2.1	110.5963	17.5372	9
967	5.5485	29.8180	0	110.5963	17.5372	9
968	5.5474	29.8040	0	110.5963	17.5372	9

Then, according to Step 4 of EWDC, we use the AVE-LNG and AVE-LAT in this new file to obtain the corresponding 18 data cube values (the temperature and conductivity) from the 18 layers of the data cube, and the partial data of the 18 data cube values are shown in Table 6. The values of three criteria under the different orders of the regression equation are shown in Table 7. The orders of the temperature and conductivity regression equations are both four. 99.06% temperature data and 99.09% conductivity data can be explained by the equation generated by the regression analysis. Thus, four order regression equations of the temperature and conductivity are reasonable and effective in this CTD file.

After determining the order of equation, we can obtain the equations about the temperature and conductivity from fitting the curve with 18 points. Next, we need to determine the thresholds of ΔT and ΔC.

Table 6 Partial data of 18 data cube values

Layer	Depth (m)	Temperature (°C)	Conductivity
1	0	29.7586	5.5219
2	10	29.6292	5.5417
3	20	29.5532	5.5494
4	30	28.6811	5.4869
…	…	…	…
15	600	7.1596	3.5193
16	700	6.2507	3.4418
17	800	5.5097	3.3815
18	1000	4.2763	3.2846

5.2 Experiments about the thresholds

In this section, we carried out experiments to find out the appropriate threshold values of ΔT and ΔC. In the new CTD file, the fitting curves of the temperature and conductivity can be obtained by four order regression equations mentioned in Section 5.1, and each depth value can derive a pair of the temperature and conductivity values which are looked as the predictive values. The fitting curve and new CTD data are shown in Fig. 9. The red curve represents the fitting curve and the blue points denote the new CTD test values.

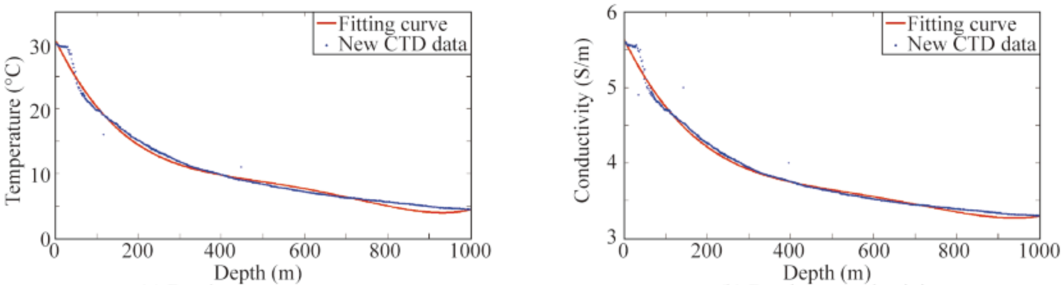
5.2.1 *Determining the temperature threshold*

We calculate the absolute values of the differences between the test values in the new CTD file and the predictive values from the fitting curve. There are 968 records in the new CTD file, so there should be 968 ΔT values.

We change the value of H_{T} to get different early warning results and find out that H_{T}= 1 is reasonable. Based on Step 6 of the EWDC, we compare all ΔT with the temperature threshold and output 34 abnormal records as the temperature warning results which are demonstrated in Fig. 10a and Table 8. The green points in Fig. 10 represent the ab-

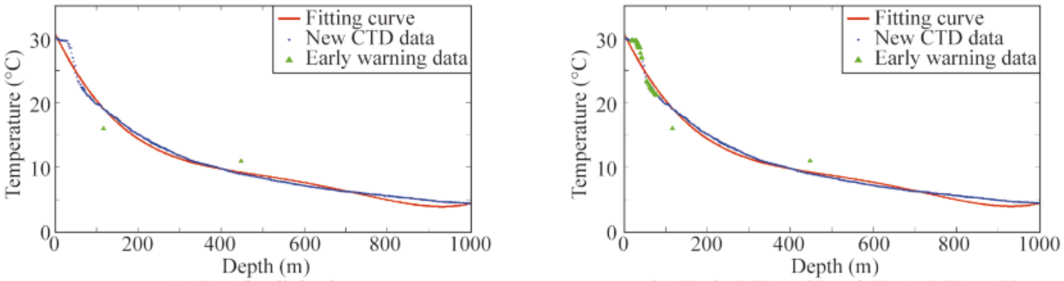
Table 7 Results of three critera

Criterion	Order								Determined order
	1	2	3	4	5	6	7	8	
$\bar{R}^2_{\mathrm{T}}$	0.7853	0.9390	0.9815	0.9906	–1.4784	–3.6122	–5.1338	–7.1011	4
AIC_{T}	2.8427	1.6300	0.4795	–0.1651	5.4454	6.0906	6.3915	6.6755	4
BIC_{T}	2.8921	1.7289	0.6279	0.0327	5.6927	6.3874	6.7378	7.0712	4
$\bar{R}^2_{\mathrm{C}}$	0.7909	0.9483	0.9853	0.9909	–9.0112	–19.7079	–28.8345	–43.9944	4
AIC_{C}	–1.9921	–3.3430	–4.5572	–5.0049	2.0335	2.7844	3.1654	3.5820	4
BIC_{C}	–1.9426	–3.2441	–4.4088	–4.8071	2.2808	3.0812	3.5116	3.9777	4



(a) Depth vs. temperature
(b) Depth vs. conductivity

Fig. 9. Fitting curve and test values.



(a) H_{T}=1 in all depths
(b) H_{T}=3, $DEP \leqslant 100$ and H_{T}=1, $DEP > 100$

Fig. 10. Early warning results of temperature.

Table 8 Detailed output of temperature early warning results

H_{T}	No.	Longitude (°E)	Latitude (°N)	Depth (m)	ΔT (°C)
H_{T}=1, $0 \leqslant DEP \leqslant 1000$	1	110.5963	17.5372	18.5	1.2506
	2	110.5963	17.5372	22.3	1.6780
	3	110.5963	17.5372	26.3	2.1380
	4	110.5963	17.5372	30.6	2.5927
	5	110.5963	17.5372	35.1	2.5990
	...	...	...	...	...
	34	110.5963	17.5372	18.0	1.1138
H_{T}=3, $0 \leqslant DEP \leqslant 100$	1	110.5963	17.5372	116.3	3.1283
H_{T}=1, $100 < DEP \leqslant 1000$	2	110.5963	17.5372	448.2	1.7748

normal values.

Considering the unstable temperature in the shallow water area, we attempt to enlarge the threshold in the shallow water area. Set H_{T}=3 in the depth of 0 m to 100 m and set H_{T}=1 in the other depth layer, the early warning results of the temperature decrease to two records as shown in Fig. 10b and Table 8. There are only two green points which apparently deviate from the fitting curve in Fig. 10b. Less green points are found when the threshold is larger from the layer of 0 m to the layer of 100 m.

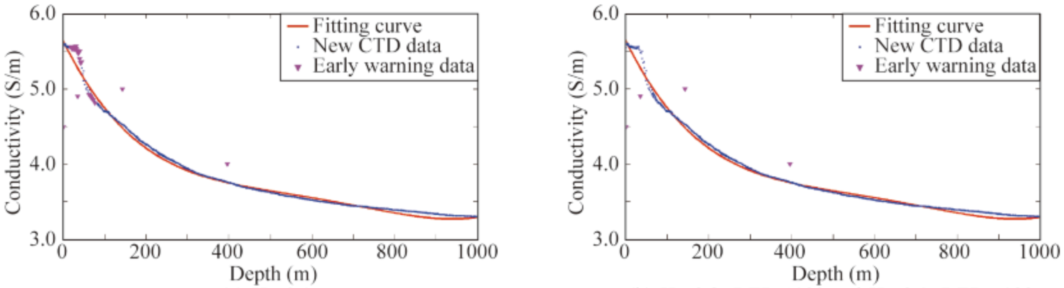
The detailed output of the temperature warning results including the longitude, latitude, depth and ΔT are presented in Table 8. We compare the two schemes and choose the latter because the unstable data in shallow water should be considered, and it will bring too many wrong early warning results if the threshold is too small in shallow water.

5.2.2 *Determining the conductivity threshold*

Similarly, we do the early warning experiments about the conductivity by changing the value of H_{C}. We find that H_{C} = 0.1 is reasonable. Fig. 11a and Table 9 show the 27 errant points filtered by the conductivity warning. Considering the instability of sampled data in shallow water, we set H_{C} = 0.3 when the depth is from 0 m to 100 m and H_{C} = 0.1 in the other depths, and the detailed conductivity warning results are shown in Fig. 11b and Table 9. As shown in Table 9, less warning results are obtained and the data which may be false alarms in Fig. 11a are deleted.

6 Conclusions

The present study was designed to fully use the marine



(a) H_C=0.1

(b) H_C=0.3, DEP≤100 and H_C=0.1, DEP>100

Fig. 11. Early warning results of the conductivity.

Table 9 Detailed output of the conductivity early warning results

H_C	No.	Longitude (°E)	Latitude (°N)	Depth (m)	ΔC (S/m)
H_C=0.1, 0≤DEP≤1000	1	110.5963	17.5372	0	1.1528
	2	110.5963	17.5372	18.5	0.1193
	3	110.5963	17.5372	22.3	0.1545
	4	110.5963	17.5372	26.3	0.1939
	5	110.5963	17.5372	30.6	0.2437
	...	...	...	...	...
	27	110.5963	17.5372	20.1	0.1131
H_C=0.3, 0≤DEP≤100	1	110.5963	17.5372	0	1.1528
	2	110.5963	17.5372	35.1	0.3776
H_C=0.1, 100<DEP≤1000	3	110.5963	17.5372	142.9	0.5202
	4	110.5963	17.5372	396.4	0.2490

data sampled by multiple underwater vehicles in real sea trails and find out the early warning method for new sampled data based on the historical data. The investigation of the sampled marine data has shown that these data have seven main characteristics, and two of the more significant characteristics are big data volume and data sparseness. These characteristics make it extremely difficult to do some meaningful applications like early warning of the marine environment based on these sampled data. This study has defined the concept of the marine data cube and designed a data fitting algorithm DFTS to construct the data cube. The data cube has demonstrated, for the first time, how to integrate the big marine data sampled along saw-tooth paths. The major contribution of this study is presenting an early warning algorithm based on the data cube (EWDC) to output the abnormal information of a new sampled data file. It is the first study to do some realistic application on the data sampled by multiple underwater vehicles. These experiments confirmed that the DFTS is effective to construct the marine data cube and the EWDC is reasonable to obtain the early warning results.

The most important limitation of this study lies in the fact that there are only one-year sampled data which only cover a limited ocean area. We cannot get the variation trend of the temperature and conductivity in the same month of different years. In the future, a large number of gliders will be deployed to carry out automatic marine monitoring in a long time. Thus, the data will be much bigger. The marine data cube construction method is good at dealing with big marine data and will serve as a foundation for future studies. The proposed algorithms will have greater application value, as well as higher early warning accuracy.

References

Båmstedt, U. and Brugel, S., 2017. A cost-precision model for marine environmental monitoring, based on time-integrated averages, *Environmental Monitoring and Assessment*, 189(7), 354.

Bouffard, J., Pascual, A., Ruiz, S., Faugère, Y. and Tintoré, J., 2010. Coastal and mesoscale dynamics characterization using altimetry and gliders: A case study in the Balearic Sea, *Journal of Geophysical Research*: *Oceans*, 115(C10), C10029.

Chai, T. and Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? *Geoscientific Model Development Discussions*, 7(1), 1525–1534.

Cheadle, C., Vawter, M.P., Freed, W.J. and Becker, K.G., 2003. Analysis of microarray data using Z score transformation, *The Journal of Molecular Diagnostics*, 5(2), 73–81.

Dobricic, S., Pinardi, N., Testor, P. and Send, U., 2010. Impact of data assimilation of glider observations in the Ionian Sea (Eastern Mediterranean), *Dynamics of Atmospheres and Oceans*, 50(1), 78–92.

Dong, L., Li, L., Li, Q.Y., Wang, H. and Zhang, C.L., 2015. Hydroclimate implications of thermocline variability in the southern South China Sea over the past 180,000 yr, *Quaternary Research*, 83(2), 370–377.

Fang, S.F., Xu, L.D., Zhu, Y.Q., Liu, Y.Q., Liu, Z.H., Pei, H., Yan, J.W. and Zhang, H.F., 2015. An integrated information system for snowmelt flood early-warning based on internet of things, *Information Systems Frontiers*, 17(2), 321–335.

Fiorelli, E., Leonard, N.E., Bhatta, P., Paley, D.A., Bachmayer, R. and Fratantoni, D.M., 2006. Multi-AUV control and adaptive sampling in Monterey Bay, *IEEE Journal of Oceanic Engineering*, 31(4), 935–948.

Jiang, P., Liu, X., Zhang, J. and Yuan, X., 2016. A framework based on hidden Markov model with adaptive weighting for microcystin forecasting and early-warning, *Decision Support Systems*, 84, 89–103.

Montgomery, D.C., Peck, E.A. and Vining, G.G., 2015. *Introduction to Linear Regression Analysis*, John Wiley & Sons, New York.

Park, S. and Lee, S.R., 2014. Red tides prediction system using fuzzy reasoning and the ensemble method, *Applied Intelligence*, 40(2), 244–255.

Park, Y., Cho, K.H., Park, J., Cha, S.M. and Kim, J.H., 2015. Development of early-warning protocol for predicting *chlorophyll-a* concentration using machine learning models in freshwater and estuarine reservoirs, Korea, *Science of The Total Environment*, 502, 31–41.

Pieri, G., Cocco, M. and Salvetti, O., 2018. A marine information system for environmental monitoring: ARGO-MIS, *Journal of Marine Science and Engineering*, 6, 15.

Qiu, C.H., Mao, H.B., Yu, J.C., Xie, Q., Wu, J.X., Lian, S.M. and Liu, Q.Y., 2015. Sea surface cooling in the northern South China Sea observed using Chinese sea-wing underwater glider measurements, *Deep Sea Research Part I*: *Oceanographic Research Papers*, 105, 111–118.

van der Aalst, W.M.P., 2013. Process cubes: Slicing, dicing, rolling up and drilling down event data for process mining, *Proceedings of the 1st Asia Pacific Conference on Business Process Management*, Springer, Beijing, pp. 1–22.

Vasilijević, A., Nađ, D., Mandić, F., Mišković, N. and Vukić, Z., 2017. Coordinated navigation of surface and underwater marine robotic vehicles for ocean sampling and environmental monitoring, *IEEE/ASME Transactions on Mechatronics*, 22(3), 1174–1184.

Xu, Z.Z., Jia, M.F., Li, L., Yu, S., Yu, J.C. and Liu, S.J., 2016. Data preprocessing and fitting algorithm based on marine data sampled by multiple underwater gliders, *Proceedings of the 12th World Congress on Intelligent Control and Automation*, IEEE, Piscataway, pp. 1036–1041.

Yu, J.C., Zhang, F.M., Zhang, A.Q., Jin, W.M. and Tian, Y., 2013. Motion parameter optimization and sensor scheduling for the sea-wing underwater glider, *IEEE Journal of Oceanic Engineering*, 38(2), 243–254.

Zhang, S.W., Yu, J.C., Zhang, A.Q. and Zhang, F.M., 2011. Steady three dimensional gliding motion of an underwater glider, *Proceedings of 2011 IEEE International Conference on Robotics and Automation*, IEEE, Shanghai, pp. 2392–2397.

Zhang, S.W., Zhang, A.Q. and Yu, J.C., 2015. Ocean observing with underwater glider in South China Seas, *Proceedings of 2015 IEEE International Conference on Cyber Technology in Automation* , Control, and Intelligent Systems, IEEE, Shenyang, pp. 1109–1114.

Zheng, G.Z., Zhu, N., Tian, Z., Chen, Y. and Sun, B.H., 2012. Application of a trapezoidal fuzzy AHP method for work safety evaluation and early warning rating of hot and humid environments, *Safety Science*, 50(2), 228–239.

Zollo, A., Lancieri, O., Lancieri, M., Wu, Y.M. and Kanamori, H., 2010. A threshold-based earthquake early warning using dense accelerometer networks, *Geophysical Journal International*, 183(2), 963–974.