

## Research Article

# Outlier Detection of Light Buoy Telemetry and Telecontrol Data Based on Improved Adaptive $\epsilon$ Neighborhood DBSCAN Clustering

Liangkun Xu,<sup>1,2</sup> Yongxing Jin ,<sup>1</sup> Han Xue ,<sup>2</sup> and Shibo Zhou<sup>2</sup>

<sup>1</sup>Merchant Marine Academy, Shanghai Maritime University, Shanghai 200000, China

<sup>2</sup>College of Navigation, Jimei University, Xiamen 361021, Fujian, China

Correspondence should be addressed to Yongxing Jin; yxjin@shmtu.edu.cn

Received 1 February 2021; Revised 2 May 2021; Accepted 15 May 2021; Published 28 May 2021

Academic Editor: Weifeng Pan

Copyright © 2021 Liangkun Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, according to the water area of light buoy, the migration rule of light buoy in main channel is counted, and the frequency of light buoy passing through a certain position point in the process of migration is calculated, and the model is verified by buoy position data. An anomaly detection algorithm based on improved adaptive DBSCAN clustering is designed. The size of the  $\epsilon$  neighborhood is adaptive according to the wind speed, wave height, and drift distance span of the water area where the light buoy is located. The experimental results show that the improved adaptive DBSCAN clustering algorithm can solve the problem that the common DBSCAN clustering algorithm takes the “hot” water area of the light buoy position or the most likely area in the light buoy migration process as the noise point.

## 1. Introduction

There are many reasons for the abnormal data of the light buoy, and the main reason is the offset and drift of the light buoy. After the light buoy drifts, the position data returned to the telemetry and telecontrol system is quite different from the normal data. In addition to the abnormal data caused by the drifts of the light buoy, the transmission and storage of the telemetry data may also cause the abnormal position data of the light buoy, from natural conditions (wind, current, and wave), external factors (ship traveling wave), and human factors (human throwing). The factors that affect the positioning accuracy of the buoy are analyzed from the following aspects: the equipment factor (satellite error) and the attitude of the buoy (the inclination angle of the buoy). The original lamp buoy telemetry data may have errors in manual input, information transmission, sensor data acquisition, storage, and other links. The analysis of abnormal data of light buoy position is an interference data for the analysis of light buoy offset. Therefore, in data processing, it is necessary to eliminate the abnormal data points of each light buoy. Therefore, it is necessary to preprocess the data before using the lamp buoy telemetry

position data, so as to provide a high-quality data set for subsequent use. The telemetry data of the light buoy includes the position, time, voltage, and other important parameters that characterize the operation state of the light buoy. The offset trajectory of light buoy is the record sequence of the position and time of the light buoy, and it is an important type of spatiotemporal data. Through the analysis of the position data, the similarity characteristics of the displacement trajectory of each light buoy can be obtained, and the meaningful offset patterns can be found. The telemetry data processing of light buoy mainly includes two aspects: one is the deletion of abnormal position points, and the other is the selection and cleaning of data after the light buoy is displaced.

Barbariol et al. discussed wave buoys on moored platforms and free floating platforms in the Southern Ocean [1]. Venkatesan et al. studied the drift of mooring buoy network in the northern Indian Ocean (Omni) at different depths below 500 meters [2]. Srinivasan et al. described a case study of indigenous drifting buoys in the Indian Ocean since 2012 [3]. Yu et al. proposed an adaptive correction mechanism of drift factor based on East China Sea shelf buoy data set [4]. Hostache et al. studied the drifting buoy with navigation

system, which can measure the water surface elevation from almost any point in the world [5].

In the aspect of using clustering algorithm for data preprocessing, Jin et al. proposed a method based on CFAR detection and density clustering to suppress the speckle interference and transverse stripe interference of ISAR image [6]. Wen et al. proposed hierarchical preprocessing of big data of sign in trajectory based on density clustering [7].

DBSCAN (density based spatial clustering of applications with noise) is a representative density based clustering algorithm. Different from partition and hierarchical clustering method, it defines cluster as the largest set of density connected points, which can divide the area with enough high density into clusters, and can find clusters of arbitrary shape in the noisy spatial database. Lin et al.'s experiments in semiconductor optical amplifier show that DBSCAN algorithm can improve the performance of 40 Gb/S16 QAM and 30 GB/S64 QAM receivers [8]. Hou et al. proposed a nonparametric clustering algorithm based on dominating set and DBSCAN algorithm [9]. Shen et al. proposed a real-time image superpixel segmentation method based on density noise and spatial clustering algorithm [10]. Bryant and Cios proposed a new density based clustering algorithm RNN-DBSCAN, which uses the reverse nearest neighbor count as the estimation of observation density [11]. Guo et al. proposed a multitarget recognition method based on prior independent density noise application spatial clustering (PI-DBSCAN) algorithm [12]. Wang and Lin proposed an improved adaptive parameter density clustering algorithm, which uses kernel density estimation to determine the reasonable interval of EPS and minPts parameters [13]. Zhang et al. proposed an image mosaic algorithm based on DBSCAN and mutual information [14].

In the area of data processing, Deng et al. proposed an improved quantum-inspired differential evolution algorithm for deep belief network [15]. Song et al. proposed a multipopulation parallel coevolutionary differential evolution for parameter optimization [16]. Deng et al. designed a differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem [17]. Song et al. designed an enhanced success history adaptive DE for parameter optimization of photovoltaic models [18]. Deng et al. proposed an enhanced MSIQDE algorithm with novel multiple strategies for global optimization problems [19].

However, DBSCAN is very sensitive to user-defined parameters. Subtle differences may lead to very different results. However, the selection of parameters is irregular and can only be determined by experience. But it is difficult to set the cluster radius and parameter threshold minPts. At present, there are few researches on the detection of big data outliers in buoy telemetry and remote control, and most of them stay in the qualitative analysis stage, while the quantitative research is less, and the research is not deep enough.

The novelty of this paper is summarized as follows: According to the position data of Xiamen Bay light buoy, this paper counts the length and width of the rectangle outside the light buoy offset position to analyze its offset range. According to the offset distance and the adaptive clustering parameters of the wind, current, and wave data in the water area where the

TABLE 1: Drift range of Xiamen Bay light buoy (unit: m).

Buoy	Latitude distance $\Delta\varphi$	Longitude distance $\Delta\lambda$
3#	78.9	68.9
4#	65.5	71.12
5#	41.11	56.67
6#	73.3	70
7#	88.9	87.78
8#	77.78	97.23
9#	96.6	100
10#	75.5	73.34
11#	84.4	90
11A#	81.1	78.9
12#	91.1	85.56
13#	88.9	96.68
14#	65.5	63.34
14A#	80.0	54.45
15#	71.6	67.7
16#	60.0	67.7
17#	68.8	73.2
18#	72.2	74.4
19#	62.2	73.3
20#	55.0	83.3
21#	47.0	66.6
22#	43.1	67.7
22A	73.3	97.7
23#	68.8	98.1
24#	44.4	77.7
28#	57.7	99.7
29#	63.3	73.3

light buoy is located, an anomaly detection algorithm based on improved adaptive DBSCAN clustering is designed. The size of the  $\epsilon$  neighborhood is adaptive according to the wind speed, wave height, and drift distance span of the water area where the light buoy is located.

## 2. DBSCAN Clustering

There are several definitions in DBSCAN:  $\epsilon$  neighborhood: the area within the radius of a given object called the neighborhood of the object;

input: database containing  $N$  points, radius  $\epsilon$ , and minimum number of points minPts; and output: all generated clusters meeting the density requirements.

- (1) Detect object  $P$  that has not been checked in the database. If  $P$  is not processed (classified as a cluster or marked as noise), check its neighborhood. If the number of objects contained is not less than minPts, establish a new cluster  $C$ , and add all the points in it to the candidate set  $N$ .
- (2) For all the unprocessed objects  $q$  in the candidate set  $n$ , the neighborhood is checked. If there are at least minPts objects, these objects are added to  $N$ ; if  $q$  is not included in any cluster,  $q$  is added to  $C$ ; if  $q$  is not included in any cluster,  $q$  is added to  $C$ .
- (3) Repeat step (2), continue to check the unprocessed objects in  $N$ , and the current candidate set  $n$  is empty.
- (4) Repeat steps (1)~(3) until all objects are clustered or labeled as noise. That is, the core point and the points

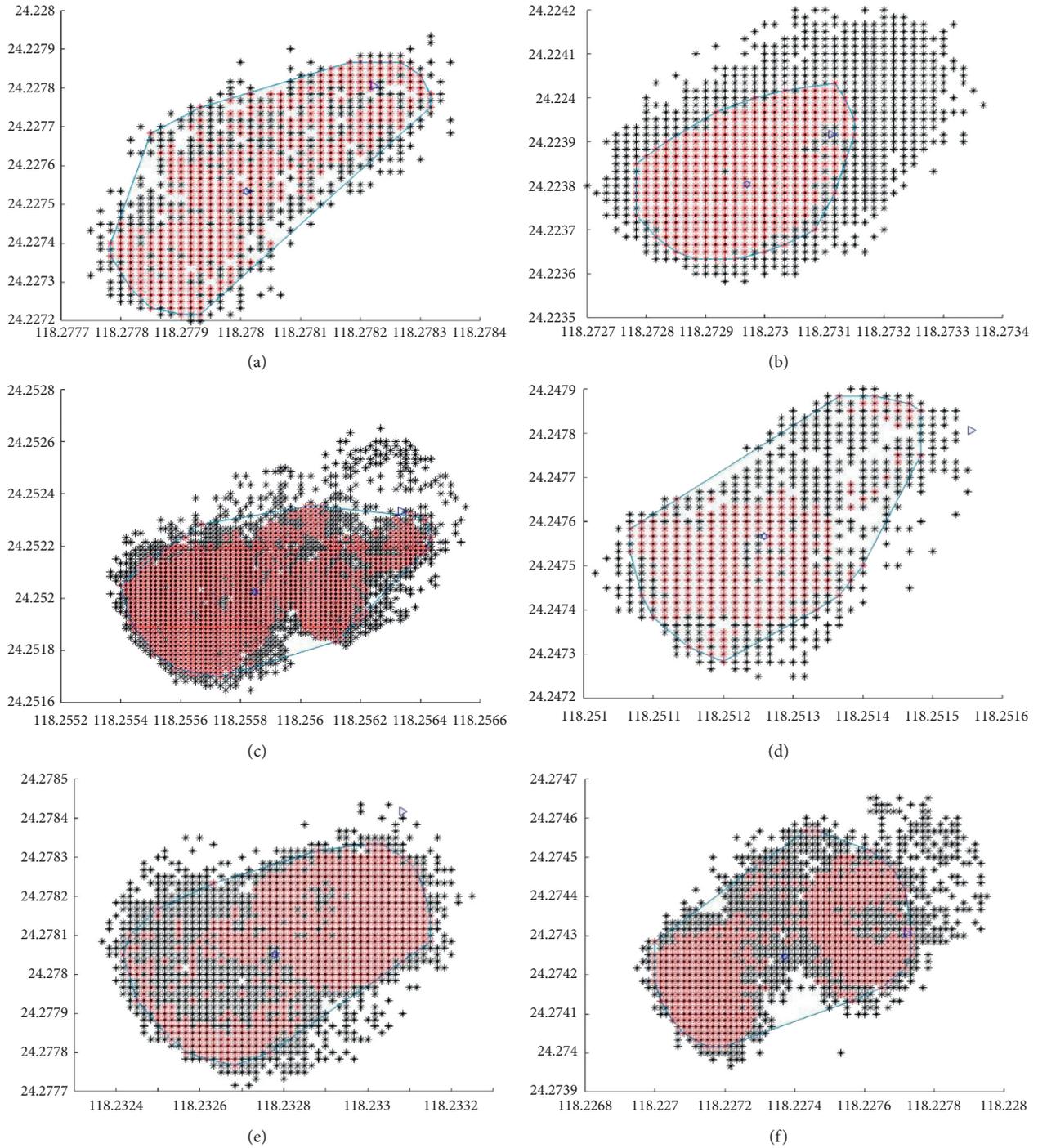


FIGURE 1: Buoy offset high-frequency position and its geometric center. (a) 3#. (b) 4#. (c) 5#. (d) 6#. (e) 7#. (f) 8#.

around it are divided into a cluster, and the boundary point is divided into the cluster of the points around it.

### 3. Statistics of the Migration Law of the Light Buoy in Xiamen Bay

According to the position data of Xiamen Bay light buoy, count the length and width of the rectangle outside the light buoy offset position to analyze its offset range. The offset values of each light buoy are shown in Table 1.

It can be seen from the statistical data that the mean value of the South-North migration range of Xiamen Bay light buoy is 69.5 m, so the mean value of South or North migration is about 35 m. The mean value of East-West migration is 78.3 m, so the mean value of East-West migration is about 39 m, which is far less than the theoretical value.

According to the empirical analysis, after several large and small tides, the anchor chain near the sinking rock is gradually silted into the seabed to form the ground chain,

which will make the actual anchor point of the light buoy shift, and the actual length of the revolving anchor chain is far less than the total length of the anchor chain, which reduces the offset range of the light buoy floating body. The direction of the resultant force of wind and current determines the direction and distance of the light buoy. The wind and current of Xiamen port have their own characteristics. In the waters near the channel, the flow direction is mostly along the channel, which is a reciprocating flow. The normal wind direction of Xiamen Bay is ENE, and the strong wind directions are SE and SW. The phenomenon of monsoon is obvious. From September to March of the next year (spring and winter), northeast East Monsoon prevails with high wind speed, while, from April to August, southeast wind prevails with low wind speed. These characteristics of wind and current determine that the light buoy may pass through a certain position frequently in the range of its active water area, and the frequency is higher. In this section, the frequency of Xiamen Bay light buoy at different position points is counted, and the center coordinates of geometric shapes of each light buoy position data set containing no less than 80% position data points are calculated. On this basis, the characteristics of light buoy offset are analyzed.

The high-frequency position points and their geometric centers of the light buoys in Xiamen Bay are shown in Figure 1. In the figure, the blue five-pointed star is the geometric center of the high-frequency position points of the light buoys, the blue triangle is the position of the sinking stone, and the blue polygon is the envelope line containing at least 80% of the position data points. The geometric center of high-frequency position points of each light buoy is shown in Table 2.

#### 4. Anomaly Detection Based on Improved Adaptive DBSCAN Clustering

This paper adopts the parameter of adaptive clustering. The size of the neighborhood is adaptive according to the wind speed, wave height, and drift distance span of the water area where the light buoy is located. The specific values are as follows:

$$\varepsilon = \left( 1 + \frac{|v_{\text{wind}}|}{|v_{\text{wind}}|_m} + \frac{|v_{\text{wave}}|}{|v_{\text{wave}}|_m} \right) \varepsilon_m, \quad (1)$$

$$\varepsilon_m = \max \left\{ \frac{\Delta\lambda}{\lambda}, \frac{\Delta\varphi}{\varphi} \right\},$$

where  $|v_{\text{wind}}|$  is the absolute value of wind speed in the water area,  $|v_{\text{wind}}|_m$  is the maximum value of wind speed in the water area,  $|v_{\text{wave}}|$  is the absolute value of wave height in the water area where the light buoy is located,  $|v_{\text{wave}}|_m$  is the maximum value of wave height in the water area where the light buoy is located,  $\Delta\varphi$  is the North-South distance span of the light buoy drift,  $\Delta\lambda$  is the East-West distance span of the light buoy drift,  $\varphi$  is the latitude, and  $\lambda$  is longitude of the light buoy drift.

API is the abbreviation of application program interface. The API of Chuanxun is a set of JavaScript functions, which

TABLE 2: Geometric center coordinates of high-frequency position points of light buoys in Xiamen Bay.

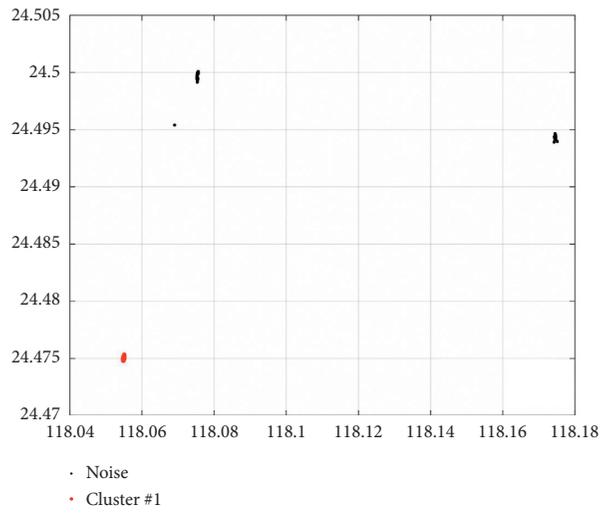
Buoy	Latitude $\varphi$ ( $^{\circ}$ ) (N)	Longitude $\lambda$ ( $^{\circ}$ ) (E)
3#	24.22753	118.2780
4#	24.2238	118.2730
5#	24.25203	118.2559
6#	24.24756	118.2513
7#	24.27805	118.2328
8#	24.27425	118.2274
9#	24.30664	118.2075
10#	24.2084	118.2036
11#	24.32106	118.1945
11A#	24.33722	118.1758
12#	24.32244	118.1784
12A#	24.33614	118.1679
13#	24.35297	118.1576
14#	24.34857	118.1533
14A#	24.35335	118.1486
15#	24.3699	118.1381
16#	24.36658	118.1336
17#	24.38412	118.1245
18#	24.38336	118.1173
19#	24.40753	118.1018
20#	24.39617	118.1048
21#	24.41279	118.0913
22#	24.40678	118.0916
22A#	24.4141	118.0794
23#	24.422	118.0753
24#	24.41822	118.0721
28#	24.42986	118.0511
29#	24.43668	118.0529

can avoid the professional and complex GIS technology and simply and conveniently embed the chart and shipping position service of Chuanxun into your own business system or website, display and manage the shipping position on the background of electronic chart, and integrate with business data. Use the third-party map service API to quickly build and obtain real-time marine meteorological data.

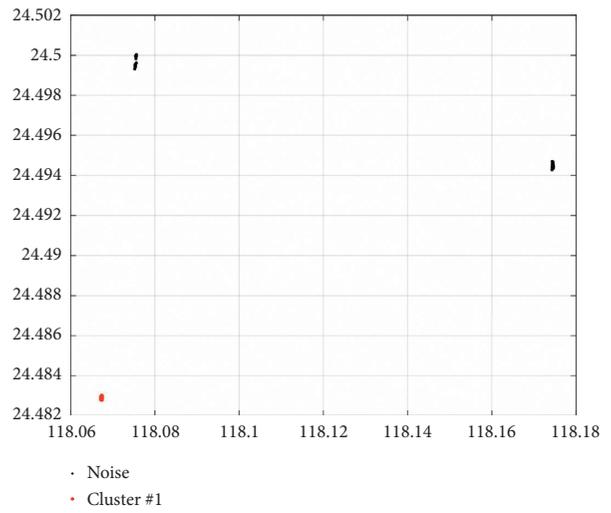
Denote  $T$  as the time needed to find the points in EPS domain.  $N$  is the number of points. The time complexity of the algorithm is  $O(NT)$ . The worst case time complexity is  $O(N^2)$ .

In low-dimensional or high-dimensional data, the spatial complexity is  $O(N)$ . For each point, it only needs to maintain a small amount of data, that is, the cluster label and the identification of each point (core point or boundary point or noise point).

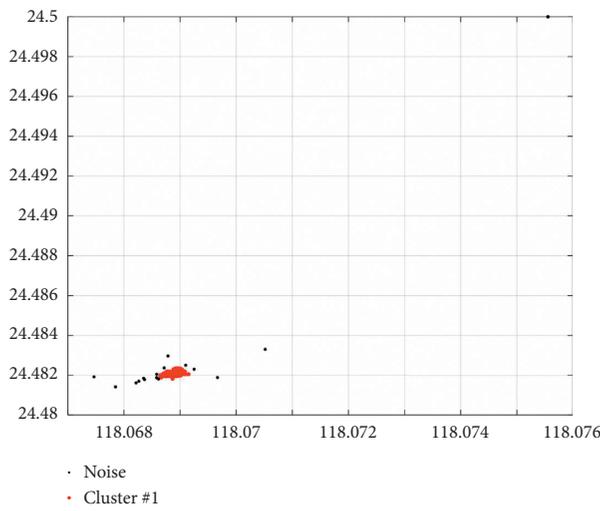
The deviations of light buoy are not equal in all directions. Under the action of wind, current, ship traveling wave, and other external forces, the deviation direction of each light buoy may have different "preferences"; that is, it will move in a small area of water in more time. The difference of light buoy migration between the inner and outer sections of Xiamen Bay is closely related to the geographical environment and flow characteristics of the port area. The sea surface in the foreign navigation section is relatively open, and the wind force in Xiamen Bay is mostly northeast wind and southwest wind. Therefore, the influence of wind on the offset of the optical buoy is not affected by the terrain



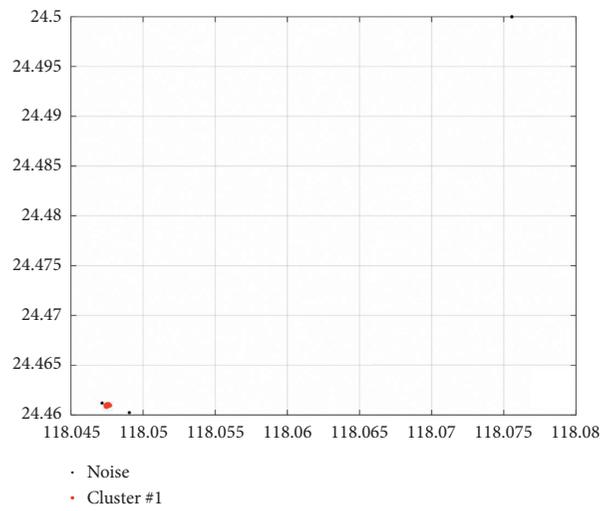
(a)



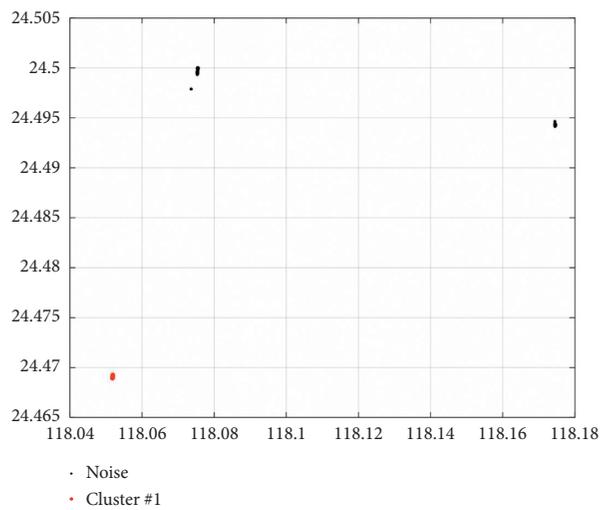
(b)



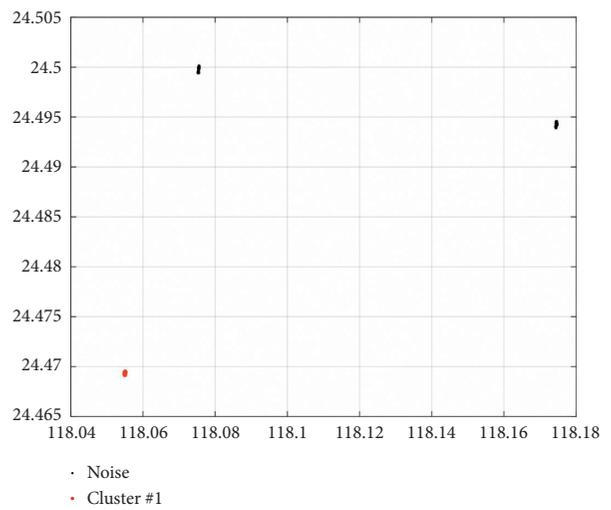
(c)



(d)



(e)



(f)

FIGURE 2: Continued.

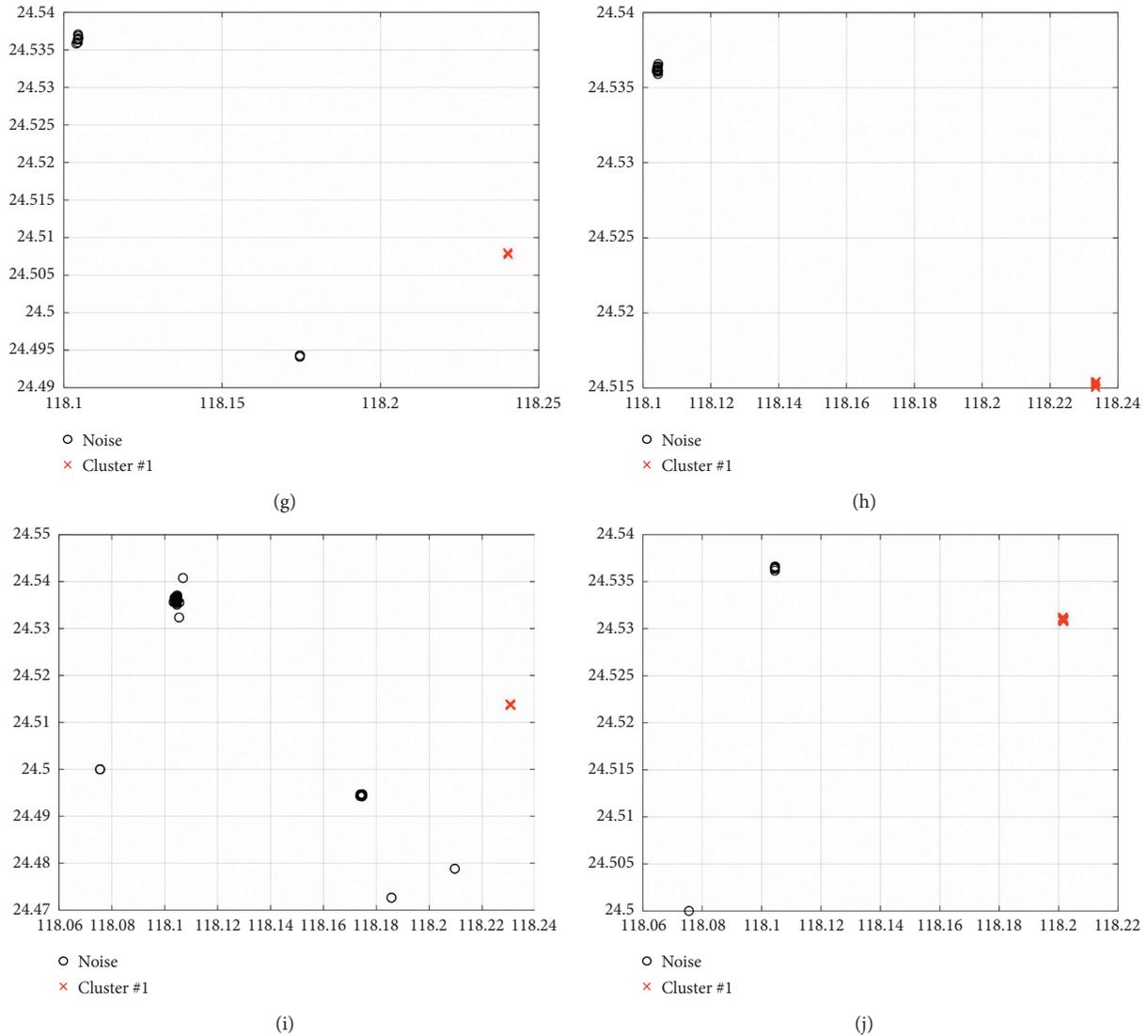


FIGURE 2: Light buoy clustering results. (a) 38#. (b) 43#. (c) 45#. (d) 34#. (e) 36#. (f) 37#. (g) JT14#. (h) JT15#. (i) JT16#. (j) JT22#.

conditions. However, for the inner leg, Xiamen Island is blocked on the north side of the #12 light buoy, and Jinmen Island is blocked on the east side. Due to the shelter of these islands, the influence of wind on buoy deviation will be weakened. This makes the influence of wind on the light buoy in the outer harbour greater than that in the inner one.

## 5. Experiment and Analysis

**5.1. Data Clustering of Light Buoy Telemetry and Telecontrol.** The deviations of light buoy are not equal in all directions. Under the action of wind, current, ship traveling wave, and other external forces, the deviation direction of each light buoy may have different preferences; that is, it will move in a small area of water in more time. Figure 2 shows light buoy clustering results.

According to the scatter diagram of the light buoy, the offset range of the light buoy and the offset rule of the light buoy under the influence of wind and current are analyzed.

TABLE 3: Reference position of Xiamen Bay light buoy.

Buoy	Anchorage offset (m)	Cyclotron radius (m)	Total length of anchor chain/m
3	38.3	9.3	
6	54.8	17.8	82.5
28	34.4	20.0	
31	49.2	12.5	
609	42	16.6	55.0

**5.2. Comparison of Different Algorithms.** Table 3 shows the reference position of the light buoy.

Table 4 shows the comparison of clustering results between ADBSCAN algorithm and basic DBSCAN algorithm. The abnormal point is considered to be the point where the distance from the sinking position exceeds the length of the anchor chain. It can be seen that, using the basic DBSCAN algorithm, sometimes the hot spot water area of the light buoy position and the most likely area in

TABLE 4: Comparison of different algorithms.

	DBSCAN (%)	ADBSCAN (%)
4#	90	96
9#	95	97
10#	96	99
11#	90	95
18#	89	93
24#	91	92
28#	93	96

the light buoy migration process are also regarded as noise points.

## 6. Conclusion

In this paper, an improved adaptive DBSCAN clustering based anomaly detection method is proposed for remote sensing data of light buoy. In the future, the clustering algorithm will be further improved, and the research results will be applied to the buoy telemetry and telecontrol data outlier cleaning.

## Data Availability

The data used to support the findings of this study are included within the article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The work was supported by Natural Science Foundation of Fujian Province (2020J01658) and the high level research and cultivation fund of transportation engineering discipline in Jimei University (HHXY2020002 and HHXY2020003).

## References

- [1] F. Barbariol, A. Benetazzo, L. Bertotti et al., "Large waves and drifting buoys in the Southern Ocean," *Ocean Engineering*, vol. 172, pp. 817–828, 2019.
- [2] R. Venkatesan, K. Ramesh, M. Arul Muthiah, K. Thirumurugan, and M. A. Atmanand, "Analysis of drift characteristic in conductivity and temperature sensors used in Moored buoy system," *Ocean Engineering*, vol. 171, pp. 151–156, 2019.
- [3] R. Srinivasan, V. Rajendran, S. Zacharia, T. Sudhakar, and M. A. Atmanand, "Indigenized Indian drifting buoys with INSAT communication for ocean observations," *Ocean Engineering*, vol. 145, pp. 263–267, 2017.
- [4] F. Yu, J. Li, Y. Zhao, Q. Li, and G. Chen, "Calibration of backward-in-time model using drifting buoys in the East China Sea," *Oceanologia*, vol. 59, no. 3, pp. 238–247, 2017.
- [5] R. Hostache, P. Matgen, L. Giustarini et al., "A drifting GPS buoy for retrieving effective riverbed bathymetry," *Journal of Hydrology*, vol. 520, pp. 397–406, 2015.
- [6] Y. Jin, C. Ye, Y. Lu, and J. Yang, "A pre-processing method of ISAR images based on CFAR detection and density clustering," *Journal of Terahertz Science and Electronic Information Technology*, vol. 18, no. 2, pp. 278–283, 2020.
- [7] R. Wen, A. Ma, X. Y. Pan, and Weiwei, "Taxonomic pre-processing of check-in trajectory big data based on density clustering," *Computer Applications and Software*, vol. 36, no. 3, pp. 20–28, 2019.
- [8] Y. Lin, E. Giacomidis, S. O. Duill, and L. P. Barry, "DBSCAN-based clustering for nonlinearity induced penalty reduction in wavelength conversion systems," *Institute of Electrical and Electronics Engineers Photonics Technology Letters*, vol. 31, no. 21, 2019.
- [9] J. Hou, H. J. Gao, and X. L. Li, "DSets-DBSCAN: a parameter-free clustering algorithm," *Institute of Electrical and Electronics Engineers Transactions On Image Processing*, vol. 25, no. 7, 2016.
- [10] J. Shen, X. Hao, Z. Liang, Y. Liu, W. Wang, and L. Shao, "Real-time superpixel segmentation by DBSCAN clustering algorithm," *Institute of Electrical and Electronics Engineers Transactions On Image Processing*, vol. 25, no. 12, pp. 5933–5942, 2016.
- [11] A. Bryant and K. Cios, "RNN-DBSCAN: a density-based clustering algorithm using reverse nearest neighbor density estimates," *Institute of Electrical and Electronics Engineers Transactions on Knowledge and Data Engineering*, vol. 30, no. 6, pp. 1109–1121, 2018.
- [12] P. C. Guo, Z. Liu, and J. J. Wang, "HRRP multi-target recognition in a beam using prior-independent DBSCAN clustering algorithm," *IET Radar, Sonar & Navigation*, vol. 13, no. 8, pp. 1366–1372, 2019.
- [13] G. Wang and G. Lin, "Improved adaptive parameter DBSCAN clustering algorithm," *Computer Engineering and Applications*, vol. 56, no. 14, pp. 45–51, 2020.
- [14] M.-Y. Zhang, Y.-Y. Wang, W. Liang-Wu, and Q. Xu-Jia, "Image mosaic algorithms based on DBSCAN clustering and mutual information," *Journal of Chinese Computer Systems*, vol. 41, no. 4, pp. 825–829, 2020.
- [15] W. Deng, H. Liu, J. Xu, H. Zhao, and Y. Song, "An improved quantum-inspired differential evolution algorithm for deep belief network," *Institute of Electrical and Electronics Engineers Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 7319–7327, 2020.
- [16] Y. Song, D. Wu, W. Deng et al., "MPPCEDE: multi-population parallel co-evolutionary differential evolution for parameter optimization," *Energy Conversion and Management*, vol. 228, Article ID 113661, 2021.
- [17] W. Deng, J. Xu, Y. Song, and H. Zhao, "Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem," *Applied Soft Computing*, vol. 100, Article ID 106724, 2021.
- [18] Y. Song, D. Wu, A. W. Mohamed, X. Zhou, B. Zhang, and W. Deng, "Enhanced success history adaptive DE for parameter optimization of photovoltaic models," *Complexity*, vol. 2021, Article ID 6660115, 22 pages, 2021.
- [19] W. Deng, J. Xu, X.-Z. Gao, and H. Zhao, "An enhanced MSIQDE algorithm with novel multiple strategies for global optimization problems," *Institute of Electrical and Electronics Engineers Transactions on Systems, Man, and Cybernetics: Systems*, vol. 2020, 10 pages, 2020.