

Article

# Ship Trajectory Clustering Based on Trajectory Resampling and Enhanced BIRCH Algorithm

Zhaojin Yan <sup>1,2,3</sup> , Guanghao Yang <sup>1,\*</sup>, Rong He <sup>4</sup>, Hui Yang <sup>1,2,\*</sup> , Hui Ci <sup>1,2</sup> and Ran Wang <sup>1,2</sup>

- <sup>1</sup> School of Resources and Geosciences, China University of Mining and Technology, Xuzhou 221116, China  
<sup>2</sup> Key Laboratory of Coal Bed Gas Resources and Forming Process of Ministry of Education, China University of Mining and Technology, Xuzhou 221116, China  
<sup>3</sup> Jiangsu Key Laboratory of Coal-Based Greenhouse Gas Control and Utilization, China University of Mining and Technology, Xuzhou 221008, China  
<sup>4</sup> Department of Civil, Environmental and Sustainable Engineering, Santa Clara University, Santa Clara, CA 95053, USA  
\* Correspondence: ts2101010a31@cumt.edu.cn (G.Y.); yanghui@cumt.edu.cn (H.Y.)

**Abstract:** Automatic identification systems (AIS) provides massive ship trajectory data for maritime traffic management, route planning, and other research. In order to explore the valuable ship traffic characteristics contained implicitly in massive AIS data, a ship trajectory clustering method based on ship trajectory resampling and enhanced BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) algorithm is proposed. The method has been tested using 764,393 AIS trajectory points of 13,845 ships in the waters of the Taiwan Strait of China, and 832 ship trajectories have been generated and clustered to obtain 172 classes of ship trajectory line clusters among 40 port pairs. The experimental results show that the proposed method has exhibited a good clustering effect on ship trajectories. Compared with the existing ship trajectory clustering methods, the proposed method can more efficiently detect and identify differences between trajectories with largely similar spatial distribution characteristics, so as to obtain legitimate clustering results. In addition, this study has constructed the main ship navigation routes between ports based on the extracted ship trajectory line clusters, and the constructed main routes are directional, refined, and rich in content compared with the existing ship routes. This research provides theoretical and technical support for ship route planning and maritime traffic management.

**Keywords:** automatic identification system (AIS) data; trajectory clustering; trajectory mining; BIRCH algorithm; main ship navigation route



**Citation:** Yan, Z.; Yang, G.; He, R.; Yang, H.; Ci, H.; Wang, R. Ship Trajectory Clustering Based on Trajectory Resampling and Enhanced BIRCH Algorithm. *J. Mar. Sci. Eng.* **2023**, *11*, 407. <https://doi.org/10.3390/jmse11020407>

Academic Editor: Claudio Ferrari

Received: 17 December 2022

Revised: 21 January 2023

Accepted: 24 January 2023

Published: 13 February 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

As the most efficient and economical mode of transporting bulk commodities over long distances, maritime transport is responsible for 90% of the world's trade flow [1]. According to the United Nations Conference on Trade and Development (UNCTAD), more than 50,000 ships sail the seas every day [2]. Therefore, ship traffic monitoring to ensure safe and smooth maritime traffic is of great significance to maritime transport; meanwhile, it is also one of the biggest challenges for maritime law enforcement, and search and rescue management has received extensive attention from researchers. Efficient and legitimate analysis of ship trajectory data can help reveal and understand ship behaviors and movement patterns, and further analysis of ship traffic flow characteristics can help identify abnormal ship behaviors, plan shipping routes, and provide valuable reference information for ship traffic monitoring.

The need for maritime traffic safety has led to the emergence of Automatic Identification Systems (AIS). AIS was originally designed to avoid ship collisions, but the growing popularity of AIS makes it possible to monitor ship trajectories on a global scale [3]. AIS, mainly composed of base station (shore-based or satellite-based) facilities and shipboard

equipment, can acquire and upload ship static information (e.g., maritime mobile service identification (MMSI) number, ship name, etc.), ship dynamic information (e.g., ship position, speed, etc.), and ship voyage information (e.g., estimated time of arrival, draught, etc.) in real time [4]. Due to its accessibility, broad coverage, and data integrity, AIS data is widely used in ship trajectory research. However, the wide spatial distribution of ships, the complexity of ship behavior, and the free route distribution make it a challenge to effectively extract useful features from the massive AIS data for research [5].

Data mining, a methodology for mining useful and potentially useful knowledge from massive data, has become the main analysis technique for massive AIS data [6]. Simply put, data mining is the process of obtaining useful, valuable, and processable data from massive data that cannot be processed by conventional methods. In the field of AIS data mining, cluster analysis is a commonly used data mining tool, which can aggregate data into different classes without a priori knowledge, interpret data of different classes, and obtain valuable information [7]. The trajectory of a ship can be regarded as a single AIS data cluster. By clustering ship trajectories, valuable information can be extracted from the seemingly chaotic AIS data, and further data analysis and comparison can be performed between similar AIS data clusters or between different classes of AIS data clusters to provide key information for ship traffic flow analysis, ship behavior classification, etc.

The Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) algorithm is a distance-based hierarchical clustering algorithm, first proposed by Zhang et al. in 1996 [8]. The BIRCH algorithm integrates hierarchical clustering and iterative repositioning methods and is suitable for processing large datasets [8]. Previous research has proved [9] that the hierarchical clustering algorithm is very effective for datasets where the number of clusters cannot be predicted in advance, and ship trajectory data happens to have this feature. Therefore, a ship trajectory clustering method based on trajectory resampling and enhanced BIRCH algorithm is proposed in this study to achieve accurate clustering of ship trajectories from massive AIS data and further extract ship routes.

The remainder of the paper is organized as follows. Section 2 reviews the current state of research on AIS data in maritime traffic, trajectory clustering, and other fields. Section 3 describes in detail the research data and research methods used in this study. Section 4 presents a case study using the data and methods in Section 3 and presents some very representative results. Then, Section 5 compares the case results with some existing classic clustering algorithms and analyzes the reasons for the anomalous clustering results of ship trajectories. Finally, Section 6 summarizes the discussion findings and limitations of this study and identifies future directions.

## 2. Literature Review

### 2.1. Research on Maritime Traffic Characteristics Based on AIS Data

The analysis of maritime traffic characteristics can provide valuable information for traffic management and planning for maritime administration departments, thereby improving maritime traffic safety and maritime navigability. The popularization of AIS equipment has enabled the use of massive AIS data. The ship speed, ship position, ship type, and other elements contained in the AIS data provide a data basis for the study of maritime traffic characteristics. Currently, AIS data has been widely used in the study of maritime traffic characteristics, such as ship traffic flow analysis [10], ship behavior identification [11], ship collision avoidance [12], fishing footprints [13], maritime trade networks [14], route planning [15], etc.

Based on AIS data, Rong et al. [12] used the Moran index and  $G_i^*$  to conduct spatial autocorrelation analysis on ship collision behaviors in the waters around Portuguese ports, and correlated the discovered hot spots with maritime traffic characteristics (average ship speed, speed dispersion, acceleration, etc.) to discover some typical traffic characteristics that led to ship collisions. Lei et al. [16] proposed an AIS-based Maritime Traffic Route Discovery (MTRD) model to discover potential ship routes, and the experimental results were consistent with the actual route distribution. Altan and Otay [17] used a grid partitioning

method based on AIS data to analyze the traffic behavior characteristics of 309,000 vessels in the Istanbul Strait in one year and uncovered potential maritime traffic hazards to help predict maritime risks. Yu et al. [18] evaluated the impact of an offshore wind farm (OWF) on offshore traffic based on AIS data collected before and after the installation of offshore wind turbines, using the Minimum Passage Distance Algorithm (MPDA) to measure the distance between ships and the OWF. The results showed that the completion of the OWF caused a change in the trajectory of passing ships, a decrease in the speed of ships, and a significant reduction in the number of passing ships.

### 2.2. Research on Ship Trajectory Clustering Based on AIS Data

Massive AIS data contains the spatial and temporal distribution rules of maritime traffic flow and ship behavior, which needs to be mined and analyzed by clustering methods. According to the different ship trajectory structures, clustering methods can be divided into point-based clustering and line-based clustering. The trajectory point-based clustering method can detect the aggregation effect of ships' spatial distribution, but it cannot discover the spatial variation pattern of the ships' behavior in time. The clustering method based on ship trajectory segments can obtain more representative characteristics of maritime traffic flow and ship behavior and is a commonly used clustering method in current research. The primary step in the line-based clustering is to calculate the similarity between ship trajectories, and the methods for calculating the distance between ship trajectory segments include structural similarity distance [19], Hausdorff distance [20], and dynamic time warping distance [21].

Zhou et al. [22] proposed a K-Means clustering method to study ship behavior based on AIS data in the inlet waters of the Port of Rotterdam, the Netherlands. It performed K-Means clustering of dynamic AIS data (speed, position, etc.) to classify different ship behaviors. The results were then processed through a Bayesian classification algorithm with static AIS data (ship's length, width, etc.) to estimate which ship behavior class the ship belongs to. The results showed that there were six ship behavior classes with great variability in spatial distribution and speed, and it was also found that ship characteristics were closely related to behavior patterns, which provided valuable references for traffic management in port areas. Rong et al. [23], based on AIS data near Portuguese port waters, implemented trajectory compression and the DBSCAN clustering algorithm to simplify ship trajectories, and a ship trajectory probability model based on vectorized ship trajectory was obtained for abnormal ship behavior detection. Wang et al. [24] also used the DBSCAN clustering algorithm to process AIS data in the surrounding waters of Australia to provide a basis for route planning and abnormal ship behavior detection. He et al. [25] extracted the turning points in AIS data in the Three Gorges Dam area and applied the DBSCAN clustering algorithm to help generate the optimal ship routes. After comparing with the actual routes, it was found that the method still has room for improvement in the face of complex waters.

### 2.3. Research in Other Fields Based on AIS Data

In recent years, with the expansion of the coverage of satellite equipment, the popularization of AIS equipment and the improvement of computer processing performance, the research based on AIS data spans many fields. In addition to improving AIS data processing and mining methods through data pre-processing [26], data compression [27], and data segmentation [5], providing AIS data services for maritime management such as illegal fishing regulation [28], and building AIS data-based maritime information service platform [29], interdisciplinary research on AIS data in the fields of environment, ecology, and economy has also emerged, such as ship emission estimation [30], fisheries ecological evaluation [13], maritime trade analysis [14], etc.

AIS equipment can provide real-time ship trajectory data in a complex maritime traffic environment, which can be used for ship trajectory prediction, risk analysis, and ship collision warning and prevention, thereby improving the safety of marine navigation [31,32].

Thanks to the high-resolution ship activity trajectories provided by AIS data, estimating ship emissions based on AIS data has become a mainstream method. Wan et al. [33] used AIS data to create the 2018 ship emission inventories of China’s Bohai Bay, Yangtze River Delta, and Pearl River Delta. Xiao et al. [30] also calculated ship emission inventories for the ports of Los Angeles and Long Beach based on 2020 AIS data and other related data. As the availability and completeness of AIS data continues to improve, detailed AIS data can also be used for ship trade statistics and analysis. Based on the AIS trajectory data of each oil tanker, Yan et al. [14] constructed a fine-grained analysis framework for global maritime oil trade in 2017, and found that the Middle East–Malacca Strait–East Asia oil shipping route is the busiest route with the largest trade volume in the world.

#### 2.4. Summary of Current Research

Most of the previous studies on ship trajectory clustering algorithms deal with all ship trajectory data in a certain region. Although it is beneficial to test the algorithm’s clustering capability in complex environments and visually present the main routes in the study area, negligence of mining minor "cold" routes fails to provide more detailed information for maritime traffic management. The current research hotspot is to improve the mainstream clustering algorithm. Cold clustering algorithms such as the BIRCH clustering algorithm outperform the mainstream clustering algorithms in some aspects, and still have the potential for in-depth research. Maritime traffic analysis, as a hotspot of AIS data research in recent years, has focused on ship behavior research based on ship trajectories. Few researchers have studied potential route mining based on ship trajectory clustering between ports. Collectively, AIS data are widely used in maritime research, and their research application areas tend to be diversified reflecting the macroscopic characteristics of maritime ship activities from the microscopic level.

Therefore, this study focuses on comparing maritime traffic characteristics between different ports, identifying different routes between ports, discovering potential routes, extracting ship trajectories between ports, and deeply exploring the hidden maritime traffic information and ship navigation movement characteristics in AIS data, so as to provide more intuitive route planning reference information for maritime traffic management.

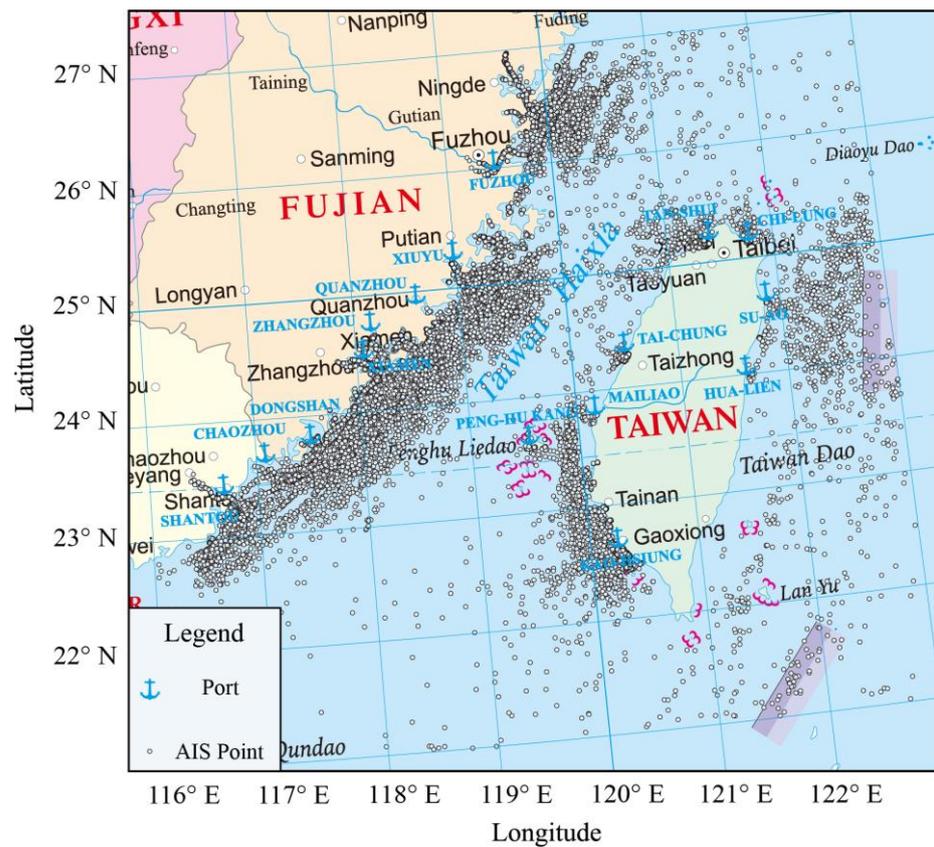
### 3. Data and Methodology

#### 3.1. Study Area and Data

The global ship AIS data collected between 116° E to 123° E and 22° N to 27° N in the Taiwan Strait region from 1 January 2017 to 31 January 2017 is shown in Figure 1. The study area contains 764,393 AIS records of 13,845 ships, and the data format is shown in Table 1. The AIS data for the study contains key attributes such as MMSI, ship speed, and ship position.

**Table 1.** Key attributes of AIS data used in this study.

Field	Significance	Example
MMSI	Maritime mobile service identification number, which is the unique identification mark of the ship.	412357870
vessel_type	Ship type, such as tankers, cargo ships, fishing vessels, etc.	Cargo ship
sog	Ship speed in knots, expressed in nautical miles per hour.	25.5
longitude	Longitude of the ship’s position in the WGS84 coordinate system, in degrees.	119.9644067° E
latitude	Latitude of the ship’s position in the WGS84 coordinate system, in degrees.	26.38369° N
utc	Coordinated universal time indicates the time when this AIS record generated.	1484102760



**Figure 1.** Study area and AIS data. Only 5% of overall AIS data is shown due to mapping effect. The background image is the standard map of the Ministry of Natural Resources of China, and the review number is GS(2020)4634.

There are 16 ports in the study area, namely Fuzhou Port, Shantou Port, Xiamen Port, Keelung Port, Su’ao Port, Hualien Port, Kaohsiung Port, Taoyuan Port, Penghu Port, Taichung Port, Mailiao Port, Chaozhou Port, Dongshan Port, Quanzhou Port, Xiuyu Port, and Zhangzhou Port, as shown in Table 2. The port data was excerpted from the World Port Index (WPI) which was published by the U.S. National Geospatial-Intelligence Agency (NGA) in 2019. WPI contains geographic information for about 3700 major ports and terminals worldwide [34].

**Table 2.** Ports in the study area.

ID	Port Name	Country	Longitude	Latitude
0	Fuzhou	CN	119.30° N	26.08° E
1	Shantou	CN	116.68° N	23.37° E
2	Xiamen	CN	118.07° N	24.45° E
3	Keelung	CN	121.77° N	25.13° E
4	Suao	CN	121.87° N	24.60° E
5	Hualien	CN	121.60° N	23.98° E
6	Kaohsiung	CN	120.25° N	22.62° E
7	Tanshut	CN	121.40° N	25.18° E
8	Penghu	CN	119.53° N	23.58° E
9	Taichung	CN	120.50° N	24.30° E
10	Mailiao	CN	120.17° N	23.78° E
11	Chaozhou	CN	117.08° N	23.62° E
12	Dongshan	CN	117.52° N	23.75° E
13	Quanzhou	CN	118.60° N	24.88° E
14	Xiuyu	CN	118.98° N	25.23° E
15	Zhangzhou	CN	118.15° N	24.68° E

### 3.2. BIRCH Clustering Method Based on Ship Trajectory Resampling

The flowchart shown in Figure 2 presents the steps to implement the BIRCH clustering method based on resampling the relevant ship AIS data.

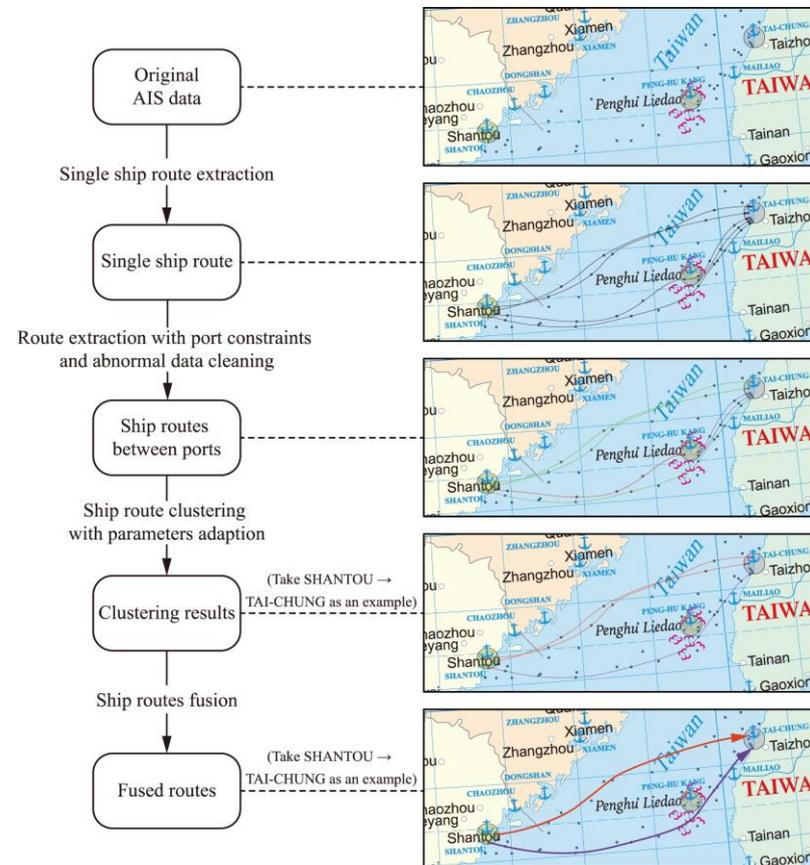


Figure 2. Flowchart of BIRCH clustering based on resampling.

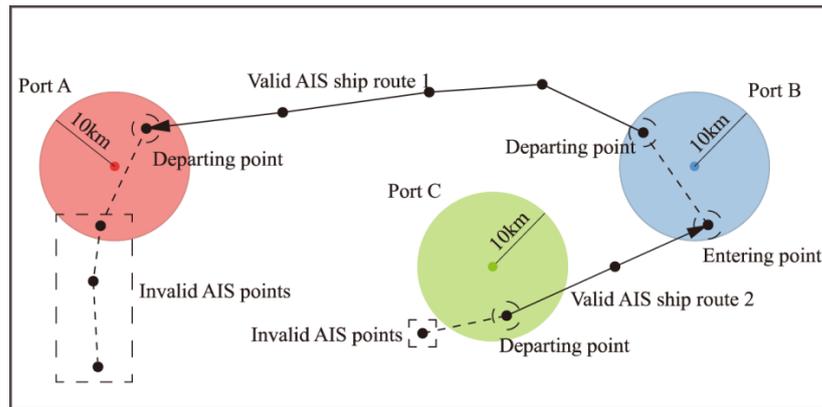
#### 3.2.1. Single Ship Trajectory Extraction

The original AIS data is a mixture of 13,845 ships’ records, which is not ready for subsequent route clustering, analysis, and fusion display. The first step is to extract the trajectory data of each ship based on its unique MMSI number. The trajectories are then coded and sorted for subsequent processing.

#### 3.2.2. Single-Ship Voyage Segmentation

The extracted single-ship trajectories are often long trajectories passing through multiple ports, so the trajectories are widely distributed in the study area. This complicates the comparison between ship trajectories and makes it difficult to analyze the traffic characteristics between ship trajectories with the same origin and destination. Therefore, in order to facilitate the study and reduce the spatial variability between ship trajectories, it is necessary to divide single-ship trajectories into single-ship voyage trajectories according to the ports they pass through. A single-ship voyage trajectory is a continuous inter-port trajectory. The specific division method is as follows.

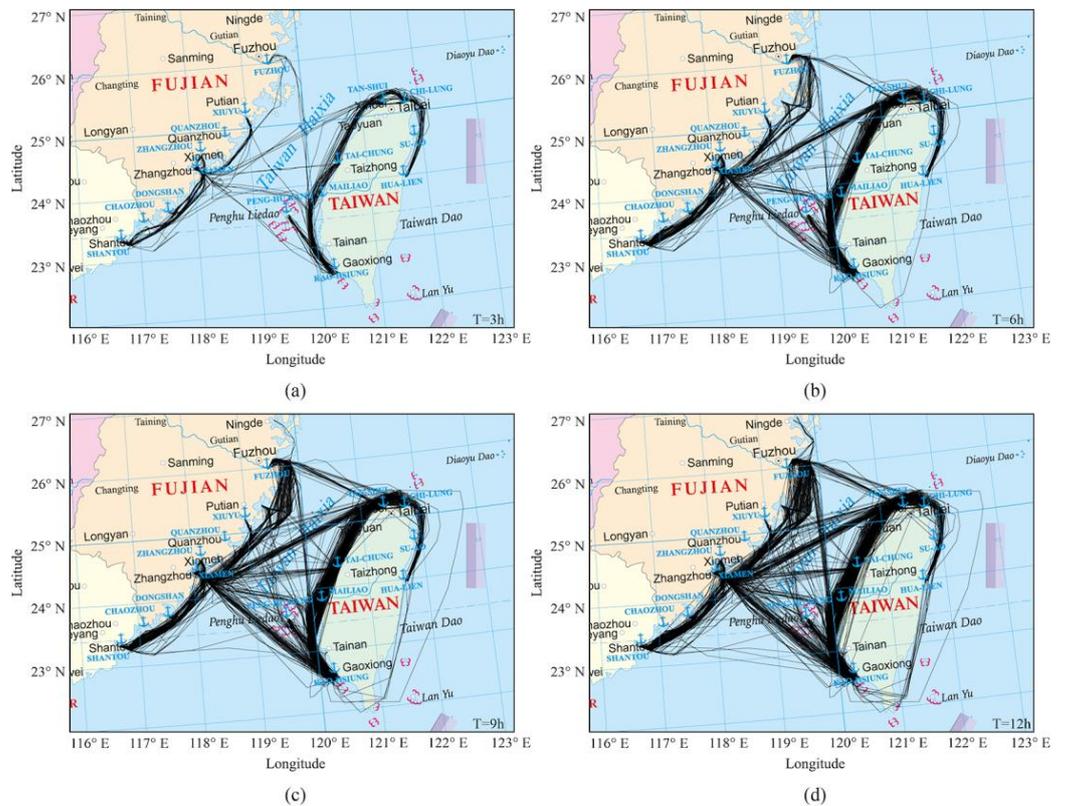
The spatial distance between each trajectory point and all ports in the study area is calculated, and if the distance is less than the specified threshold, the point is appointed as a “port point”. If there are two or more “port points” in the route, the route between the two “port points” will be intercepted, coded and sorted, and the origin and destination ports will be recorded. Since the average port width in the study area is about 10 km and the distance between ports is much larger than 10 km, the threshold is set to 10 km. The division process is shown in Figure 3.



**Figure 3.** Schematic diagram of single-ship voyage division based on inbound and outbound identification.

### 3.2.3. Data Cleaning

The time difference between consecutive ship trajectory points varies, and extreme time differences will lead to abnormal spatial distribution of ship trajectories, which is not conducive to the subsequent ship trajectory clustering. Figure 4 shows the extracted ship trajectories between ports with time difference thresholds ( $T$ ) of 3 h, 6 h, 9 h, and 12 h. It can be seen that almost no erroneous trajectories cross land when  $T$  is 6 h, while the number of erroneous trajectories increases as  $T$  increases to 9 and 12 h. It was also observed that a large number of ship trajectories are missed when  $T$  is 3 h. Therefore, the time difference threshold is 6 h and the ship trajectory data with time difference greater than 6 h were regarded as abnormal and deleted.



**Figure 4.** Ship trajectories obtained with different time difference thresholds between consecutive trajectory points. (a–d) correspond to the time difference thresholds of 3 h, 6 h, 9 h, and 12 h, respectively.

### 3.2.4. Adaptive Parameter Clustering of Ship Trajectories

To discover and compare the differences in ship trajectories between different port pairs, the clustering process is divided into five parts: the construction of similarity matrix based on Hausdorff distance, the establishment of ship trajectory clustering evaluation method, the selection of ship trajectory resampling value, the adaptive BIRCH clustering of ship trajectories, and the calculation of ship trajectory fusion. For the convenience of illustration, the ship trajectories from Kaohsiung Port to Xiamen Port are taken as an example to explain the following steps, as shown in Figure 5.

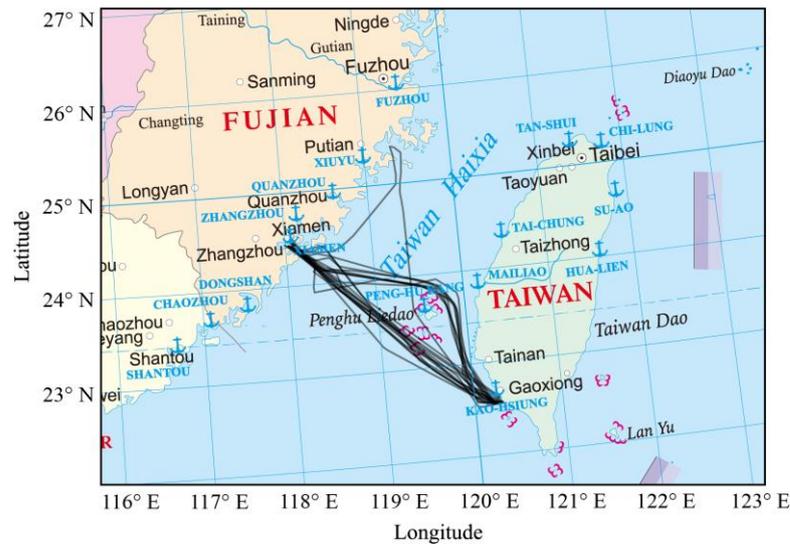


Figure 5. Ship trajectory map from Kaohsiung Port to Xiamen Port.

#### Similarity Matrix Construction Based on Hausdorff Distance

The evaluation metrics in the ship trajectory clustering process need to be established based on a certain similarity measure, hence this study constructs the similarity matrix based on Hausdorff distance and similarity function.

For two ship trajectories  $traj_A = (a_1, a_2, \dots, a_i, \dots, a_n)$  and  $traj_B = (b_1, b_2, \dots, b_i, \dots, b_n)$ , the Hausdorff distance is [35]:

$$\begin{aligned}
 H(traj_A, traj_B) &= \max\{h(traj_A, traj_B), h(traj_B, traj_A)\} \\
 h(traj_A, traj_B) &= \max\{\min\{\|a_i - b_j\|\}\} \\
 h(traj_B, traj_A) &= \max\{\min\{\|b_i - a_j\|\}\}
 \end{aligned}
 \tag{1}$$

where  $\|\cdot\|$  is the Euclidean distance between the points of  $traj_A$  and  $traj_B$ . The specific schematic diagram is shown in Figure 6.

After calculation, the distance matrix composed of the Hausdorff distances between ship trajectories from Kaohsiung port to Xiamen port is shown in Table 3 To facilitate the measurement of the similarity between ship trajectories in the Hausdorff distance matrix, the following formula is used to convert the Hausdorff distance matrix into a similarity matrix. The individual similarity is calculated as follows,

$$A_{i,j} = \begin{cases} 1, & 0 < i = j \leq n \\ S_{i,j}, & 0 < i \neq j \leq n \end{cases}
 \tag{2}$$

$$S_{i,j} = e^{-\frac{d(i,j)^2}{2\sigma_i\sigma_j}}$$

where  $\sigma_i$  is the Hausdorff distance from ship trajectory  $i$  to ship trajectory  $j$ , and  $\sigma_j$  is the average Hausdorff distance between other ship trajectories.  $A_{i,j}$  is the final similarity between ship trajectories  $i$  and  $j$ . The final similarity matrix is shown in Table 4.

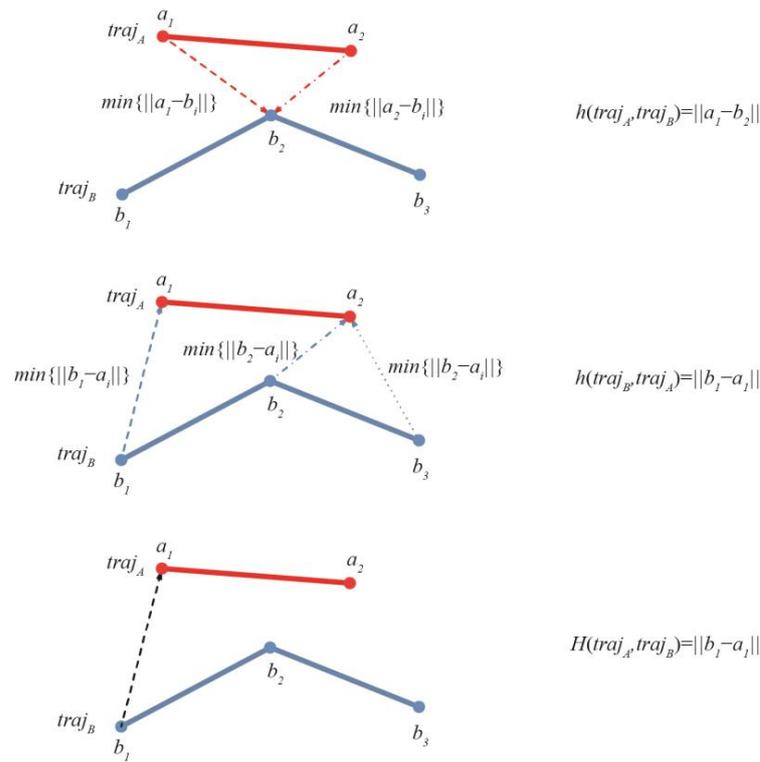


Figure 6. Schematic diagram of Hausdorff distance.

Table 3. Hausdorff distance matrix of ship trajectories from Kaohsiung Port to Xiamen Port.

Ship Trajectory ID	0	1	2	3	4	5	6	...	22
0	0	82.67	96.72	97.86	90.29	97.33	91.21	...	88.33
1	82.67	0	99.11	87.99	100.82	113.42	84.96	...	40.75
2	96.72	99.11	0	21.08	26.55	17.46	117.42	...	115.28
3	97.86	87.99	21.08	0	33.31	37.56	117.53	...	125.09
...	...	...	...	...	...	...	...	...	...
22	88.33	40.75	115.28	125.09	122.69	109.8	81.72	...	0

Table 4. Similarity matrix of ship trajectories from Kaohsiung Port to Xiamen Port.

Ship Trajectory ID	0	1	2	3	4	5	6	...	22
0	1	0.607	0.449	0.439	0.494	0.463	0.576	...	0.587
1	0.607	1	0.448	0.530	0.432	0.368	0.633	...	0.897
2	0.449	0.448	1	0.958	0.934	0.973	0.360	...	0.362
3	0.439	0.530	0.958	1	0.898	0.879	0.358	...	0.301
...	...	...	...	...	...	...	...	...	...
22	0.587	0.897	0.362	0.301	0.313	0.417	0.673	...	1

### Quantitative Evaluation Indices of Ship Trajectory Clustering Effect

To objectively evaluate the effectiveness of ship trajectory clustering, the Silhouette Coefficient (SC) [36], Davies Bouldin Index (DBI) [37], and Comprehensive Clustering Performance Metrics (CCPM) [35] were used for a comprehensive evaluation of the clustering results. SC can measure how similar a sample is to its genus clusters compared to other clusters. The SC value is in the range of  $[-1, 1]$ . The larger the value, the better the matching degree between the sample and its genus clusters than the neighboring clusters, that is, the better the clustering effect, as calculated in Equation (3) where  $p(x)$  represents the average distance between samples within a cluster, and  $q(x)$  represents the minimum average distance from the sample to other clusters.

$$SC = \sum_{i=1}^n \frac{p(x_i) - q(x_i)}{\max\{p(x_i), q(x_i)\}} / n \tag{3}$$

DBI is used to measure the distance between samples within one cluster, and a smaller DBI means a better clustering effect. The DBI value is calculated by Equation (4) where  $S_i$  and  $S_j$  represent the average distances between samples within two given clusters, and  $M_{ij}$  represents the distance between the centroids of the two clusters.

$$DBI = \sum_{i=1}^n \frac{S_i + S_j}{M_{ij}} / n \tag{4}$$

CCPM is a combination of SC and DBI, and its calculation formula is as follows:

$$CCPM = SC + 1/DBI \tag{5}$$

According to Equation (5), larger CCPM means a better clustering effect. CCPM will be mainly used as the basis for evaluating the clustering effect in the subsequent study.

### Ship Trajectory Resampling and Evaluation

The ship trajectory resampling value will affect the clustering effect and clustering speed of the BIRCH clustering algorithm, so it is necessary to find the most suitable ship trajectory resampling value. By traversing the ship trajectory resampling threshold from 5 to 100, the BIRCH clustering algorithm was performed on the trajectory data and the average SC, DBI, and CCPM are calculated as shown in Table 5. It can be seen that the most suitable resampling value is 15, so the subsequent BIRCH clustering will use 15 as the resampling value.

**Table 5.** Quantitative evaluation of ship trajectory clustering from Kaohsiung Port to Xiamen Port with different resampling thresholds.

Resampling Value	Average Cluster Number	Average DBI	Average SC	Average CCPM
5	4.990886	1.750512	-0.01113	1.05724
6	4.990886	1.750512	-0.01113	1.05724
7	4.990886	1.750512	-0.01113	1.05724
8	4.990886	1.750512	-0.01113	1.05724
9	4.990886	1.750512	-0.01113	1.05724
10	4.738936	1.770629	-0.00365	1.06221
11	5.222826	1.725781	-0.01795	1.063864
12	5.222826	1.725781	-0.01795	1.063864
13	5.222826	1.725781	-0.01795	1.063864
14	5.222826	1.725781	-0.01795	1.063864
15	5.211191	1.721574	-0.01756	1.073352
16	5.222826	1.725781	-0.01795	1.063864
17	5.222826	1.725781	-0.01795	1.063864
18	5.222826	1.725781	-0.01795	1.063864
19	5.222826	1.725781	-0.01795	1.063864
20	4.713233	1.773299	-0.00289	1.061604
25	4.713233	1.773299	-0.00289	1.061604
30	4.713233	1.773299	-0.00289	1.061604
35	4.713233	1.773299	-0.00289	1.061604
40	4.72555	1.778805	-0.00332	1.049663
45	4.72555	1.778805	-0.00332	1.049663
50	4.72555	1.778805	-0.00332	1.049663
55	4.72555	1.778805	-0.00332	1.049663
60	4.72555	1.778805	-0.00332	1.049663
65	4.72555	1.778805	-0.00332	1.049663
70	4.72555	1.778805	-0.00332	1.049663
75	4.72555	1.778805	-0.00332	1.049663
80	4.72555	1.778805	-0.00332	1.049663
85	4.72555	1.778805	-0.00332	1.049663
90	4.72555	1.778805	-0.00332	1.049663
95	4.72555	1.778805	-0.00332	1.049663
100	4.75124	1.776065	-0.00407	1.050397

### Determination of Adaptive Parameters for Ship Trajectory BIRCH Clustering

The BIRCH clustering algorithm requires two parameters: one is the maximum sample radius threshold  $T$  for each clustering feature of the leaf nodes, which determines the radius threshold of the hypersphere formed by all samples for each clustering feature; the other is the number of clusters  $N$ , which determines the number of clusters in the final clustering result [38]. The parameter selection ranges are  $T \in [0.001, 2]$  and  $N \in [2, \eta]$ , where the minimum radius interval is 0.001 and  $\eta$  is the number of ship trajectories between ports. The selection ranges of  $T$  and  $N$  were traversed in turn in the ship trajectory clustering process, and three evaluation indices, SC, DBI, and CCPM, were computed to evaluate the clustering results. Once CCPM achieves the maximum value,  $T$  and  $N$  are recorded as the final BIRCH clustering parameters.

### 3.2.5. Ship Trajectory Fusion Calculation

By clustering ship trajectories between port pairs, potential routes can be discovered and explored.

A virtual ship trajectory can be obtained as the representative route of a cluster by fusing the trajectory data belonging to the same cluster between a port pair. The specific way is to use the ship trajectory resampling method described in the previous section to resample all ship trajectories with the same sampling value to obtain  $n$  equally spaced data sets. The longitude and latitude coordinates of the matched  $n$  data sets are averaged to obtain the fused virtual ship trajectory (Figure 7).

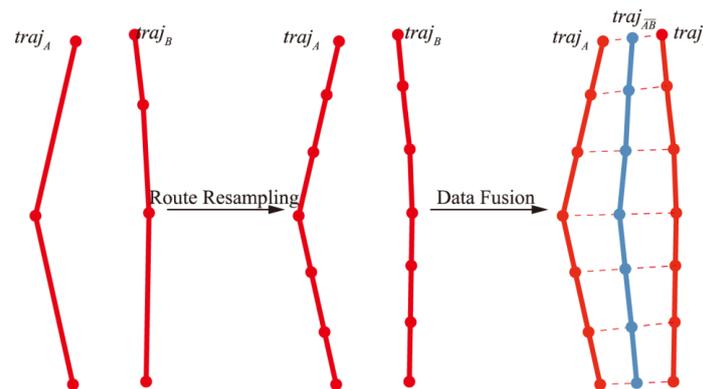


Figure 7. Ship trajectory fusion extraction.

## 4. Results

The 764,393 AIS trajectory points of 13,845 ships in the study area generated 832 ship trajectories (Figure 8a), and the enhanced BIRCH clustering resulted in 172 ship trajectory clusters between 40 port pairs (Figure 8b).

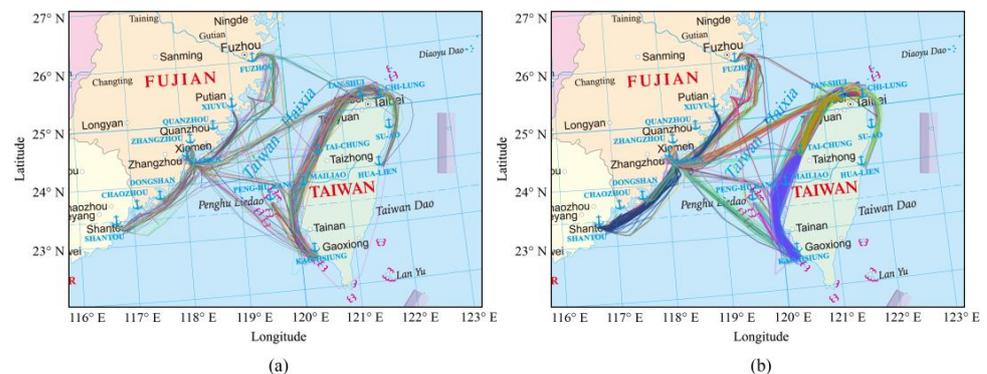


Figure 8. Spatial distribution of ship trajectories and trajectory clustering results between ports in the study area. (a) shows the ship trajectories extracted based on AIS data. (b) shows the clustering result of ship trajectories, where trajectories with the same color belong to the same clustering class.

#### 4.1. Clustering Results and Evaluation of Ship Trajectories between Port Pairs

Figure 9a–h shows the clustering results of ship trajectories between different port pairs. It can be seen that most ship trajectories between port pairs have obtained good clustering results. The computed SC, DBI, and CCPM indicators showed that CCPM has the most stable performance with obvious peaks (Figure 9a–d). However, some poor clustering results were observed, manifested as excessive number of clusters, or even pseudo-clusters (Figure 9e–h). At the same time, the CCPM evaluation of ship trajectory clustering results showed that the peak is not unique. The CCPM spikes occurred when the number of clusters was small and large. The planning of ship routes between port pairs needs to consider a variety of complex factors, such as fuel consumption, cost, distance, time, safety, etc. [5]. Therefore, the main routes between each port pair should be fixed and not too many. Based on this, it can be seen that the excessive number of ship trajectory clusters between port pairs is an anomalous result, and some adjustments to the clustering algorithm need to be made later. The specific solution will be elaborated in the discussion section.

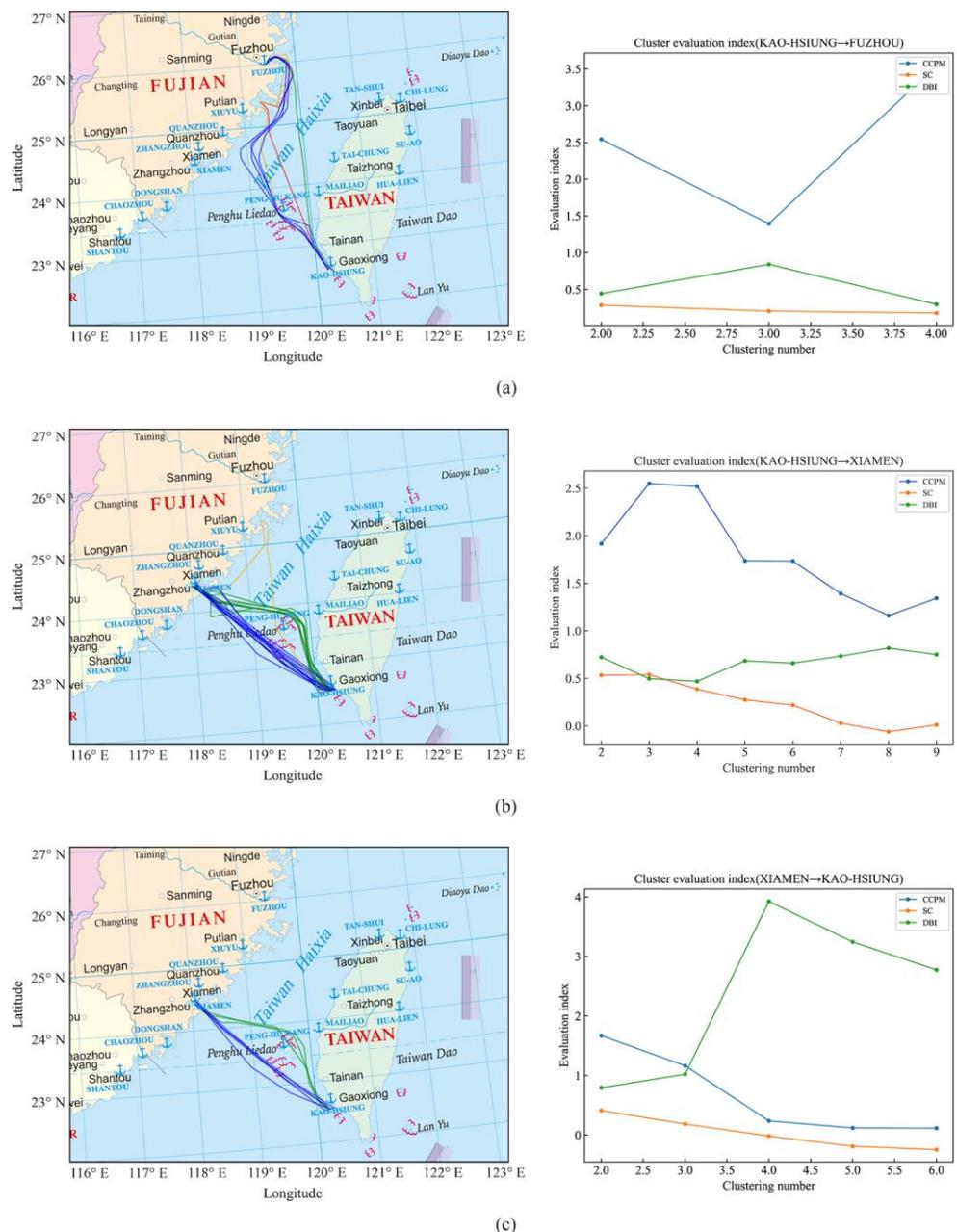


Figure 9. Cont.

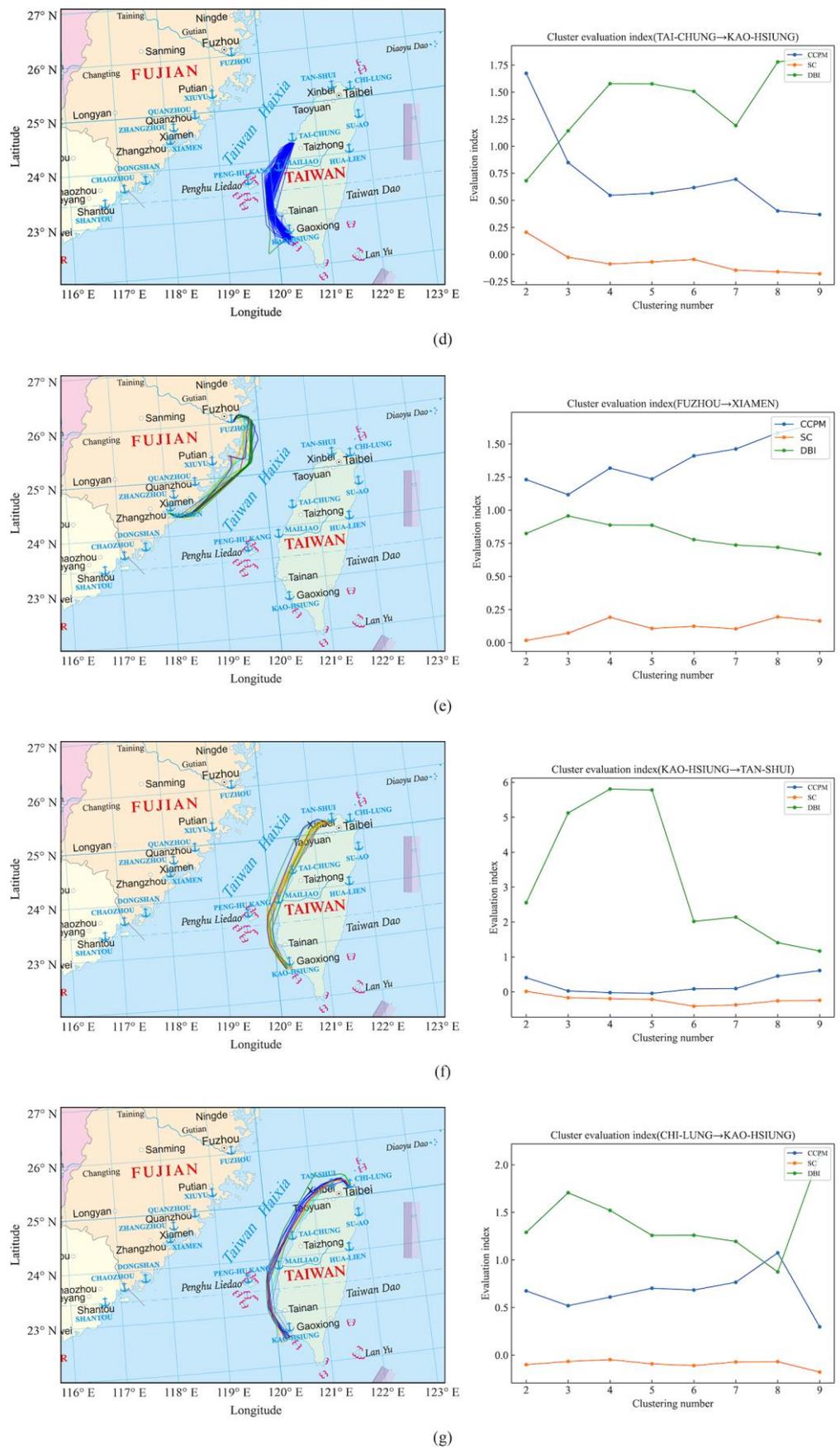
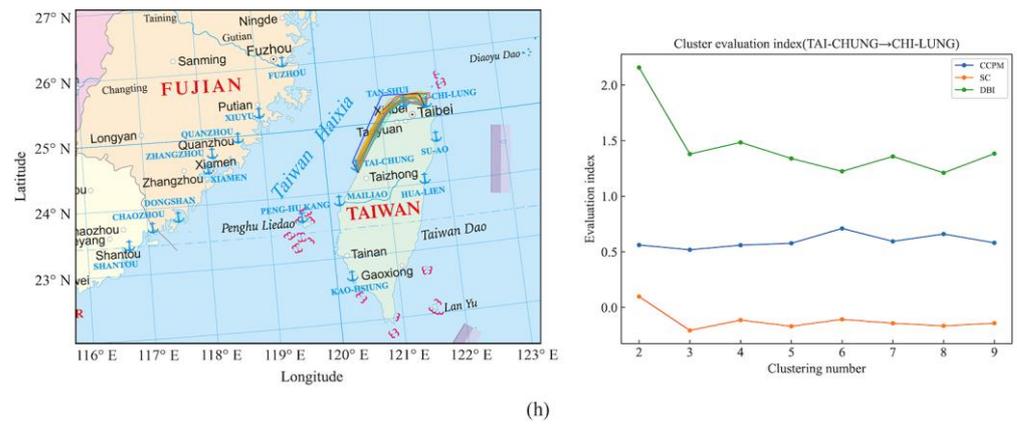


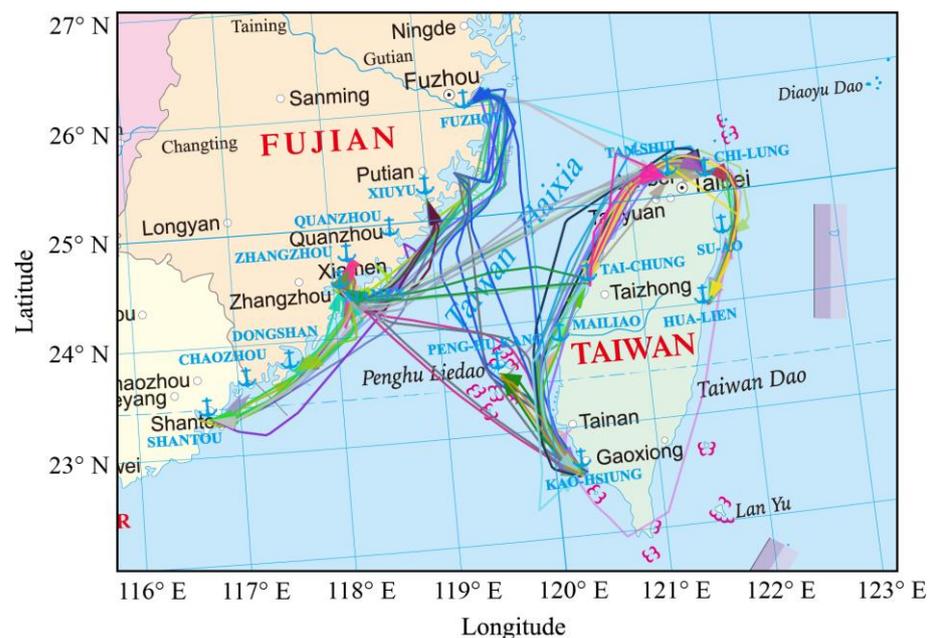
Figure 9. Cont.



**Figure 9.** (a) Clustering results of ship trajectories from Kaohsiung Port to Fuzhou Port, and four groups of ship trajectory clusters were obtained. (b) Clustering results of ship trajectories from Kaohsiung Port to Xiamen Port, and three groups of ship trajectory clusters were obtained. (c) Clustering results of ship trajectories from Xiamen Port to Kaohsiung Port, and two groups of ship trajectory clusters were obtained. (d) Clustering results of ship trajectories from Taichung Port to Kaohsiung Port, and two groups of ship trajectory clusters were obtained. (e) Clustering results of ship trajectories from Fuzhou Port to Xiamen Port, and nine groups of ship trajectory clusters were obtained. (f) Clustering results of ship trajectories from Taoyuan Port to Kaohsiung Port, and nine groups of ship trajectory clusters were obtained. (g) Clustering results of ship trajectories from Keelung Port to Kaohsiung Port, and eight groups of ship trajectory clusters were obtained. (h) Clustering results of ship trajectories from Taichung Port to Keelung Port, and six groups of ship trajectory clusters were obtained.

4.2. Main Route Extraction in the Taiwan Strait

The main ship navigation routes between ports were extracted by ship trajectory fusion, and Figure 10 shows the results in the study area. The extraction results were compared with the Ocean Passages for the World (OPW) [39]. Figure 11 shows some routes in the Taiwan Strait included in the OPW map. The maps in Figure 11 were superimposed to obtain a complete route map in the Taiwan Strait. The overlay results are shown in Figure 12.



**Figure 10.** Extraction results of the ship navigation main routes in the Taiwan Strait.

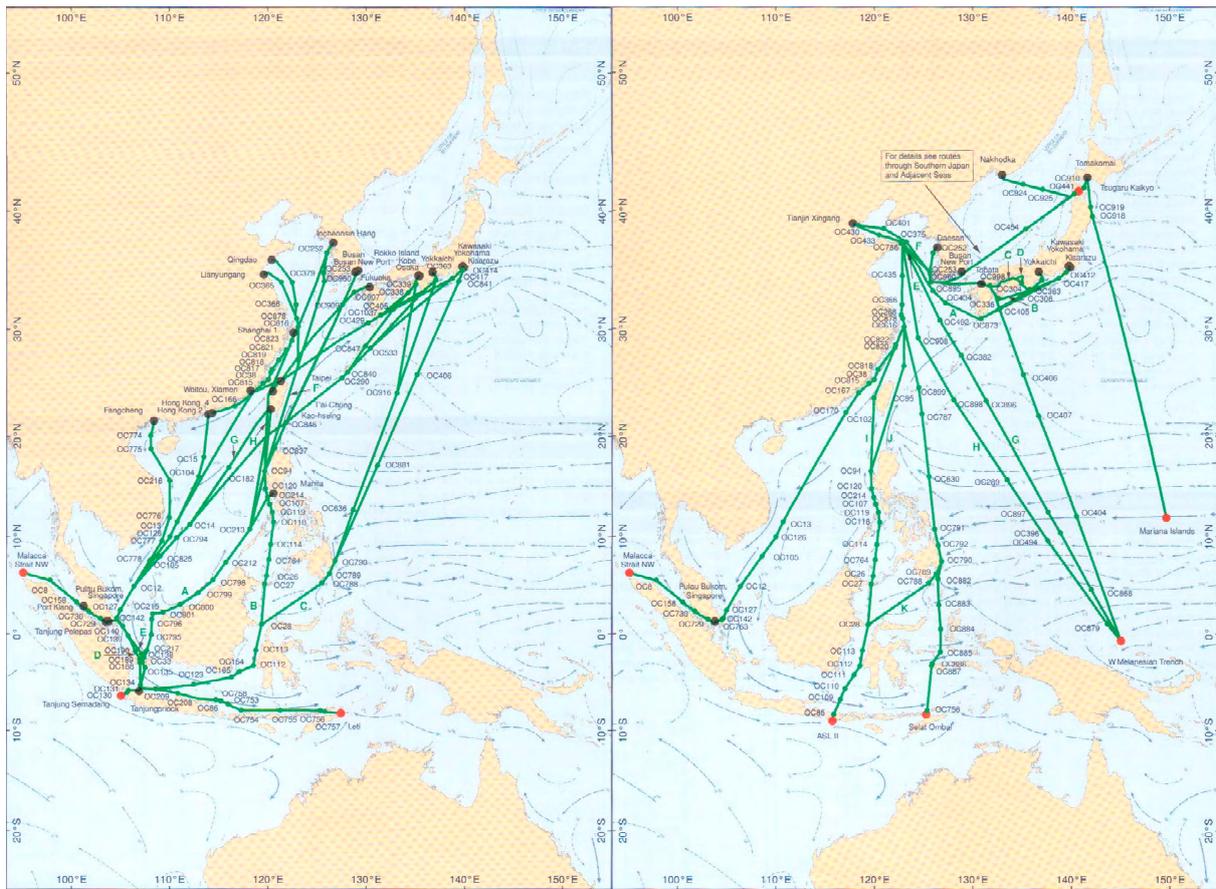


Figure 11. Some routes in the Taiwan Strait shown in the Ocean Passages for the World.

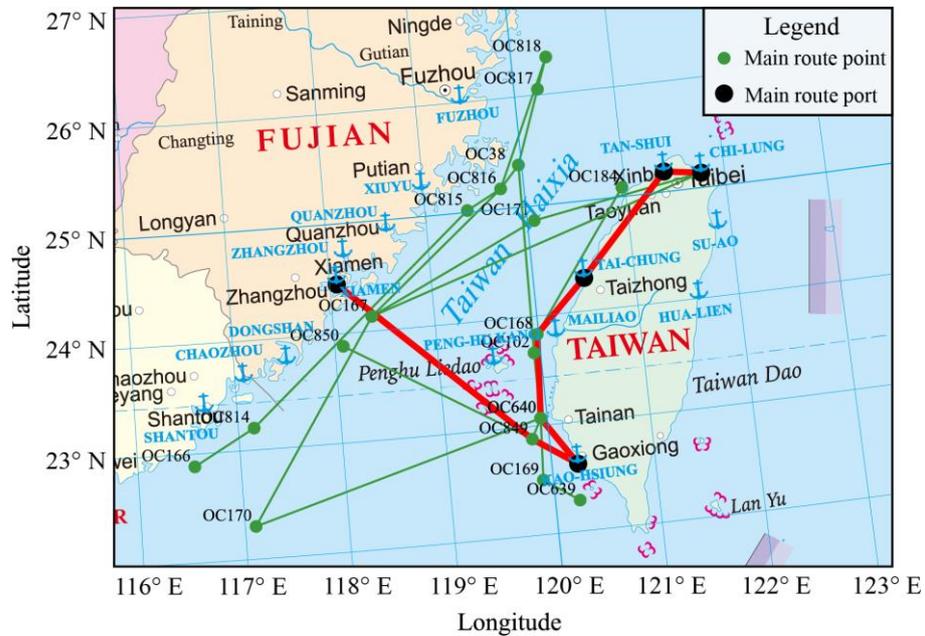


Figure 12. Route map of Taiwan Strait based on the Ocean Passages for the World.

Comparing Figure 10 with Figure 12, it can be found that the main ship navigation routes in the Taiwan Strait in the OPW route map only include the Xiamen Port–Kaohsiung Port–Taichung Port–Taoyuan Port–Keelung Port routes, while other routes are outside the study area. In addition, the OPW routes are not directional. Although the main inter-

port routes extracted in this study were similar to the OPW routes in terms of overall spatial distribution, the extracted routes were better than the OPW routes in terms of detail richness and the number of inter-port routes. At the same time, the extracted routes were directional, containing subtle differences in spatial distribution. In summary, the main route extraction based on ship trajectory resampling and the BIRCH clustering can obtain detailed spatial distribution of main routes between port pairs based on AIS data. The results were directional, detailed and rich in content, providing valuable reference for maritime traffic management and route planning.

### 5. Discussion

#### 5.1. Comparison between the BIRCH Clustering Algorithm and Mainstream Clustering Algorithms

In order to verify the difference of clustering effect between the enhanced BIRCH algorithm and mainstream ship trajectory clustering algorithms, the K-Means algorithm [40] and the DBSCAN algorithm [41] were selected as counterexamples. Table 6 gives the descriptions of the three clustering algorithms. The K-Means algorithm, DBSCAN algorithm, and enhanced BIRCH algorithm were compared under the same conditions (i.e., the input data are all ship trajectories in the Taiwan Strait waters) for ship trajectory clustering and evaluation, and the SC, DBI, and CCPM evaluation indices are calculated for each port pair. The results are shown in Figures 13–15.

Table 6. Description of the three clustering algorithms.

Clustering Algorithm	Description
K-Means	A distance-based iterative algorithm for cluster analysis requires a pre-specified value of K [40].
DBSCAN	A density-based clustering algorithm that finds arbitrarily-shaped clusters in a noisy spatial database [41].
Enhanced BIRCH	A distance-based hierarchical clustering algorithm that first uses a bottom-up hierarchical algorithm and then improves the results by iterative repositioning [38].

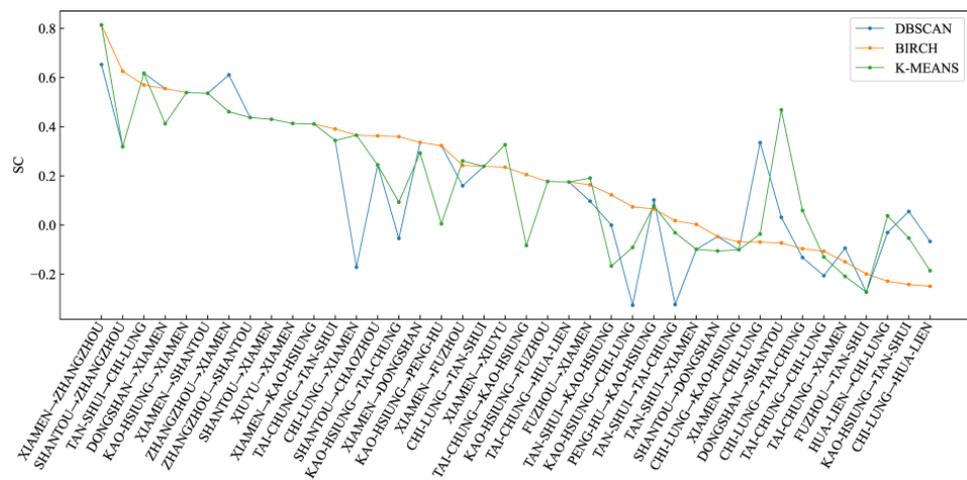


Figure 13. Comparison of SC evaluation indices for different clustering algorithms.

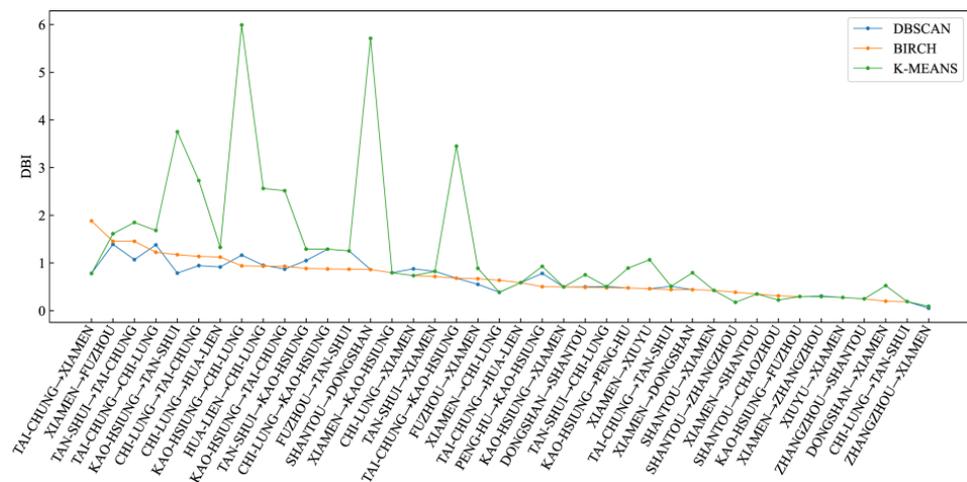


Figure 14. Comparison of DBI evaluation indices for different clustering algorithms.

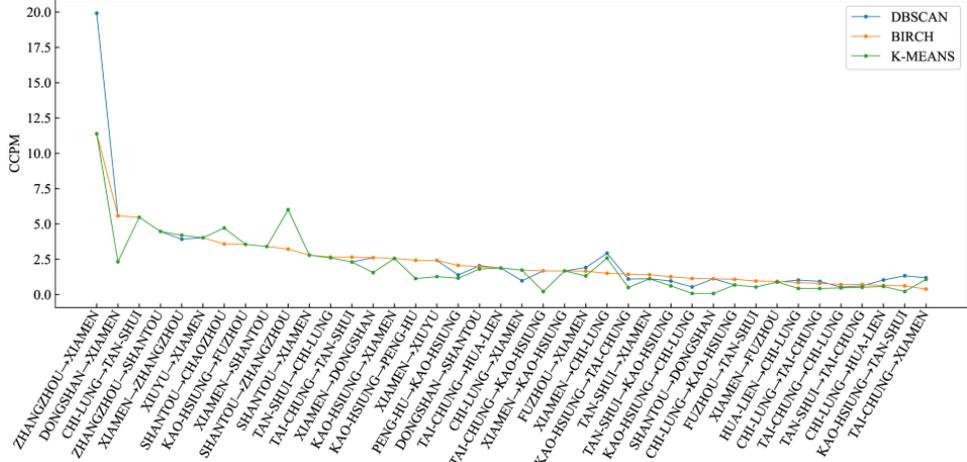
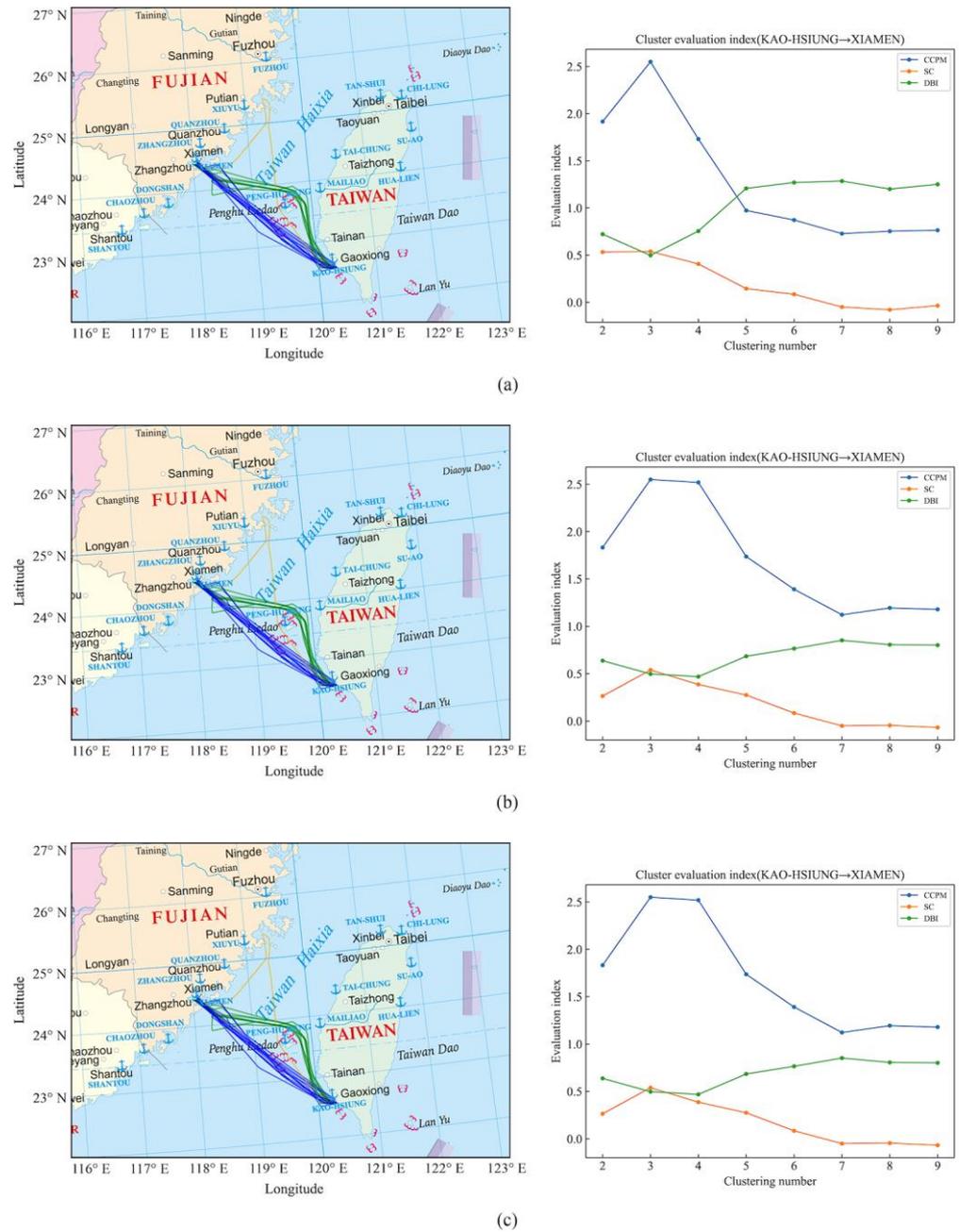


Figure 15. Comparison of CCPM evaluation indices for different clustering algorithms.

It can be seen that the enhanced BIRCH clustering algorithm was more stable than the other two clustering algorithms under the same port pair conditions, while all three evaluation metrics (SC, DBI, and CCPM) show that the BIRCH clustering algorithm scored higher than the other two clustering algorithms for most port pairs. The evaluation metrics of the K-Means and DBSCAN clustering algorithms were more volatile than the BIRCH clustering algorithm, indicating their clustering quality is questionable.

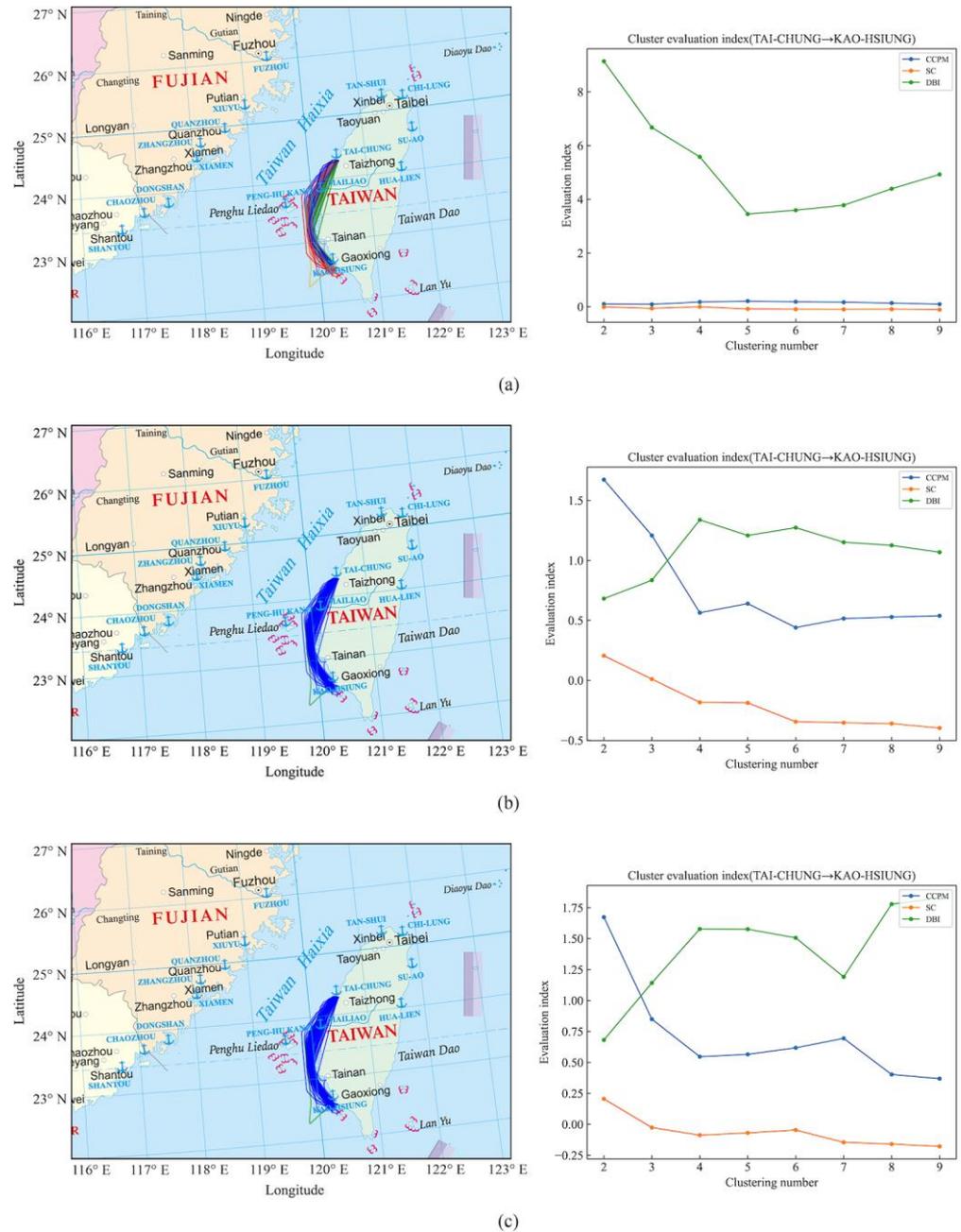
Taking the ship trajectories with big differences in spatial distribution from Kaohsiung Port to Xiamen Port as an example, the three clustering algorithms (K-Means, DBSCAN, and BIRCH) were used to cluster the ship trajectories, and the evaluation results are shown in Figure 16a–c. It can be seen that all three clustering algorithms gave similar and quality clustering results and the fluctuations and peaks of each evaluation index were obvious, which demonstrates that for ship trajectories with big differences in spatial distribution, all three clustering methods are competent.



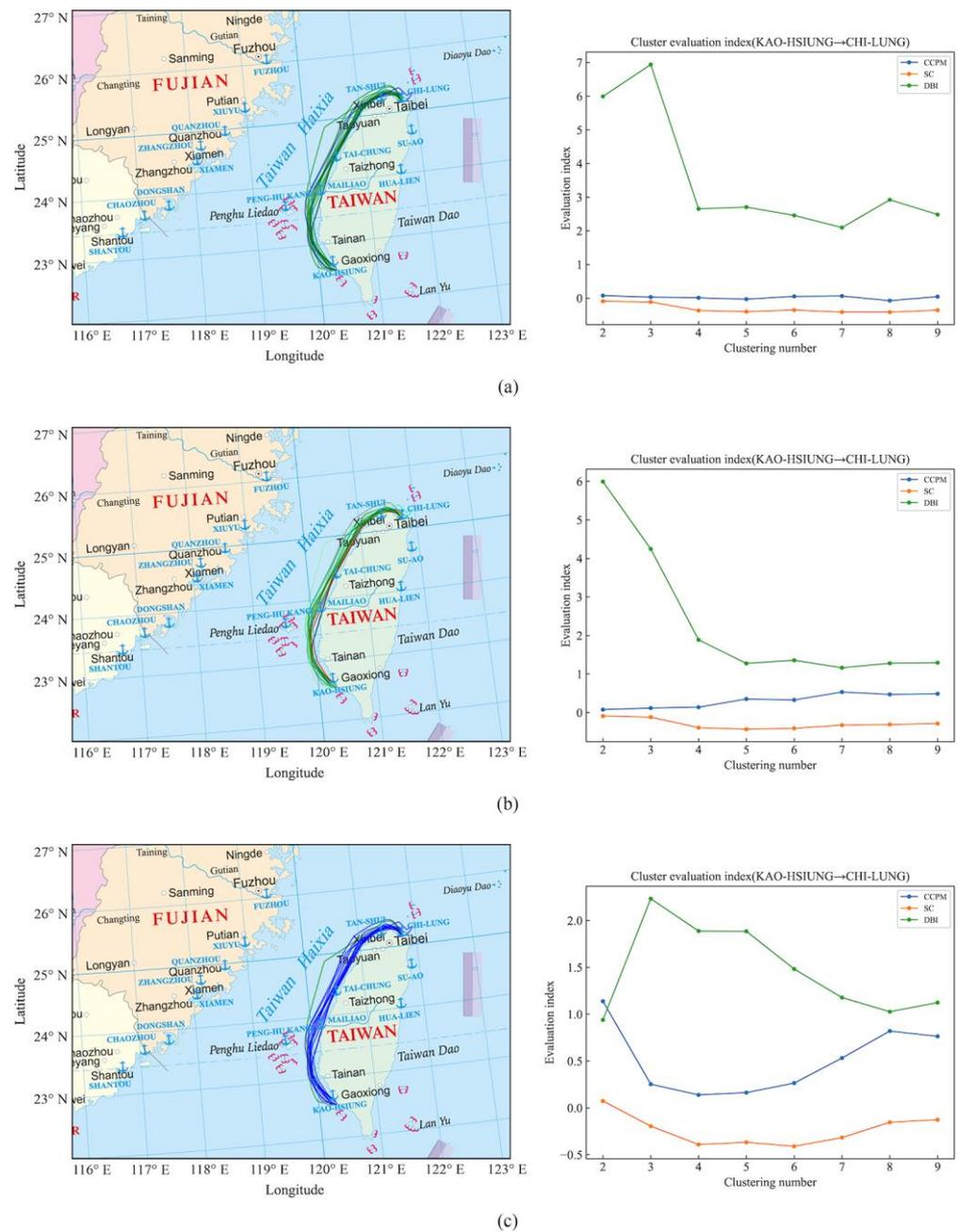
**Figure 16.** (a) K-Means clustering results and quantitative evaluation results of the ship trajectories from Kaohsiung Port to Xiamen Port. (b) DBSCAN clustering results and quantitative evaluation results of the ship trajectories from Kaohsiung Port to Xiamen Port. (c) BIRCH clustering results and quantitative evaluation results of the ship trajectories from Kaohsiung Port to Xiamen Port.

For ship trajectories with similar spatial distribution, such as those from Taichung port to Kaohsiung port (Figure 17) and from Kaohsiung Port to Keelung Port (Figure 18), the clustering effects and cluster evaluation indices of the three clustering algorithms exhibited great differences. Specifically, the K-Means clustering algorithm could not effectively distinguish between ship trajectories that are overly similar, and the SC evaluation index was relatively flat without peaks, resulting in too many classes after clustering (Figures 17a and 18a). Compared to the K-Means clustering algorithm, the DBSCAN clustering algorithm could provide more reasonable clustering results for the ship trajectories from Taichung Port to Kaohsiung Port (Figure 17b). However, when dealing with the ship trajectory data from Kaohsiung Port to Keelung Port, the DBSCAN clustering results

showed a flat trend in the evaluation indices and could not effectively cluster trajectories (Figure 18b). Both the K-Means and DBSCAN clustering algorithms could not effectively cluster ship trajectories with very similar spatial distribution characteristics. In contrast, the enhanced BIRCH clustering algorithm gave effective clustering results for ship trajectories with similar spatial distribution characteristics (Figures 17c and 18c), and its clustering evaluation indices showed obvious peaks, indicating great differences among clusters.



**Figure 17.** (a) K-Means clustering results and quantitative evaluation results of the ship trajectories from Taichung Port to Kaohsiung Port. (b) DBSCAN clustering results and quantitative evaluation results of the ship trajectories from Taichung Port to Kaohsiung Port. (c) BIRCH clustering results and quantitative evaluation results of the ship trajectories from Taichung Port to Kaohsiung Port.



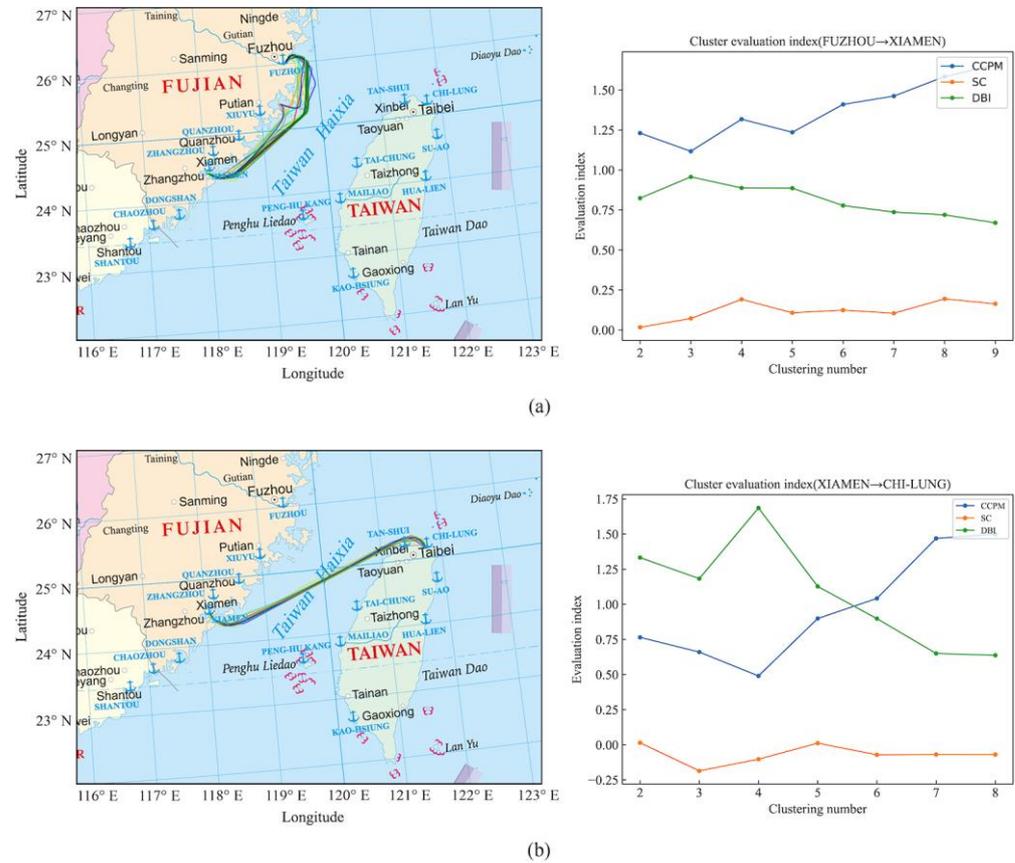
**Figure 18.** (a) K-Means clustering results and quantitative evaluation results of the ship trajectories from Kaohsiung Port to Keelung Port. (b) DBSCAN clustering results and quantitative evaluation results of the ship trajectories from Kaohsiung Port to Keelung Port. (c) BIRCH clustering results and quantitative evaluation results of the ship trajectories from Kaohsiung Port to Keelung Port.

In summary, for both ship trajectories with obvious differences in spatial distribution and ship trajectories with relatively similar spatial distributions, the enhanced BIRCH clustering algorithm can effectively provide reasonable clustering results, and its clustering performance is more stable and better than that of the K-Means and DBSCAN clustering algorithms.

### 5.2. Processing Abnormal Clustering Results Based on BIRCH Clustering Algorithm

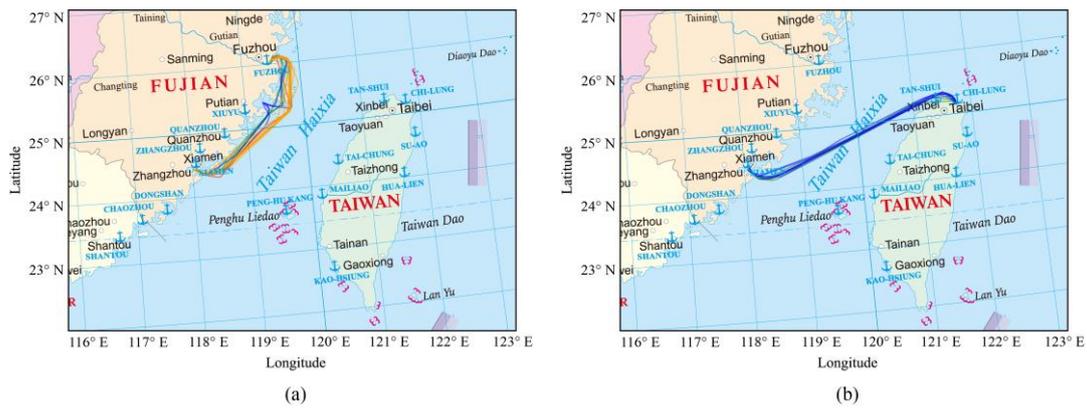
Taking the ship trajectories from Fuzhou Port to Xiamen Port (Figure 19a) and from Xiamen Port to Keelung Port (Figure 19b) as examples, the clustering results obtained from the BIRCH clustering algorithm showed pseudo-clustering, i.e., too many classes were

classified. After checking the quantitative evaluation indices of the clustering results, it can be concluded that the final number of clusters obtained from the maximum CCPM evaluation index was too large, making it impossible to get reasonable clustering results.



**Figure 19.** (a) BIRCH clustering results and quantitative evaluation results of the ship trajectories from Fuzhou Port to Xiamen Port. (b) BIRCH clustering results and quantitative evaluation results of the ship trajectories from Xiamen Port to Keelung Port.

Compared to the K-Means and DBSCAN clustering algorithms, the BIRCH clustering algorithm had greater volatility and tended to exhibit more peaks in every quantitative evaluation index when clustering ship trajectories, which indicates that the abnormal clustering results are not caused by the performance limitations of the algorithm. It is observed that there exists an extreme CCPM1 value between the clustering numbers 2 and 4 and an extreme CCPM2 value after the clustering number 5 (Figures 18c and 19a,b). When  $CCPM1 \geq CCPM2$ , the ship trajectory clustering results are reasonable and valid (Figure 19). When  $CCPM1 < CCPM2$ , the final number of clustering classes becomes too large, which leads to abnormal ship trajectory clustering results (Figure 19a,b). Therefore, based on historical observation and actual conditions, the maximum number of main routes between port pairs is specified to be five, and the extreme values of CCPM are discarded. Figure 20 shows the modified ship trajectory clustering results obtained after applying the above rules, and it can be seen that the anomalies of pseudo-clustering were effectively suppressed.



**Figure 20.** The modified clustering results of ship trajectories in Figure 20 based on the enhanced BIRCH algorithm. (a) shows the clustering results of ship trajectories from Fuzhou Port to Xiamen Port, where the ship trajectories are clustered into four groups. (b) shows the clustering results of ship trajectories from Xiamen Port to Keelung Port, where the ship trajectories are clustered into two groups.

**6. Conclusions**

Maritime traffic and ship behavior features are contained implicitly in massive AIS data. In order to further explore the valuable information in AIS data to support maritime traffic management, a ship trajectory clustering method based on ship trajectory resampling and enhanced BIRCH algorithm was proposed. The method was tested using AIS data from the waters of the Taiwan Strait of China. The 764,393 AIS trajectory points of 13,845 ships in the study area generated 832 ship trajectories, and the clustering resulted in 172 ship trajectory clusters between 40 port pairs.

Based on the clustering results of ship trajectories between port pairs, the data fusion of ship trajectories was carried out to generate main navigation routes through trajectory resampling. Comparison of the generated main routes with the documented OPW routes showed that the extracted main routes and the OPW routes have a high degree of overlap in the overall spatial distribution, while the extracted main routes were much better than the OPW routes in both richness and fineness. More importantly, the main routes extracted in this study have the directionality of ship navigation between ports, which is not available in the OPW routes. In summary, the main routes of ship navigation extracted in this study are beneficial to provide decision-making reference for maritime traffic management and route planning.

Two classical clustering algorithms, K-Means and DBSCAN, were used to compare with the enhanced BIRCH clustering algorithm to verify the effectiveness of the proposed ship trajectory clustering method. The results showed that the BIRCH clustering algorithm has advantages over the other two algorithms because its clustering results were more reasonable and effective and the CCPM evaluation index values were higher and more stable. When dealing with ship trajectories with similar spatial distribution characteristics, the BIRCH clustering algorithm can still distinguish the subtle differences and thus obtain better clustering results.

The research method in this study still has some limitations. Compared with the other two classic clustering algorithms, the BIRCH clustering algorithm is affected by multiple sets of parameters, requires more clustering within the range of parameter variation, and consumes longer time than the other two algorithms. The sample size of ship trajectory resampling has not been adapted according to the ship trajectory data of a specific port, which affects the performance and efficiency of the clustering algorithm. Therefore, in the next stage, we will seek the dynamic selection of multi-parameter ranges to reduce the time consumption of the BIRCH clustering algorithm and investigate more intelligent resampling parameter selection.

The proposed ship trajectory clustering method based on ship trajectory resampling and enhanced BIRCH clustering algorithm provides a new perspective for mining ship navigation routes from massive AIS data. The fine-grained ship navigation routes extracted in this study display the navigation conditions of different routes in complex waters in an intuitive way. The navigation conditions of different inter-port routes, such as average speed, navigation distance, navigation time, ship type, safety evaluation, etc., will be analyzed in the future study, so as to provide more granular knowledge support for maritime traffic management and route planning.

**Author Contributions:** Conceptualization, Z.Y.; Methodology, Z.Y. and G.Y.; Writing—original draft, Z.Y., G.Y. and R.H.; Writing—review & editing, H.Y. and H.C.; Formal analysis, R.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (42201451), the China Postdoctoral Science Foundation (2022M723379), the Third Comprehensive Scientific Investigation Project of Xinjiang (2022xjkk1006); the Xinjiang Uygur Autonomous Region Key Research and Development Program (2022B01012-1); the Open Research Fund of Jiangsu Key Laboratory of Coal-based Greenhouse Gas Control and Utilization, China (2022KF05), and the Fundamental Research Funds for the Central Universities (2022QN1058).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The AIS data used to support the results of this study were provided by Tianjin Seaview Technology Corporation under license and so cannot be made freely available.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Yan, Z.; He, R.; Yang, H. The small world of global marine crude oil trade based on crude oil tanker flows. *Reg. Stud. Mar. Sci.* **2022**, *51*, 102215. [[CrossRef](#)]
2. Unctad. *Review of Maritime Transport 2019*; United Nation Publication: New York, NY, USA; Geneva, Switzerland, 2019.
3. Cheng, L.; Yan, Z.; Xiao, Y.; Chen, Y.; Zhang, F.; Li, M. Using big data to track marine oil transportation along the 21st-century Maritime Silk Road. *Sci. China Technol. Sci.* **2019**, *62*, 677–686. [[CrossRef](#)]
4. Harati-Mokhtari, A.; Wall, A.; Brooks, P.; Wang, J. Automatic Identification System (AIS): Data Reliability and Human Error Implications. *J. Navig.* **2007**, *60*, 373. [[CrossRef](#)]
5. Yan, Z.; Xiao, Y.; Cheng, L.; He, R.; Ruan, X.; Zhou, X.; Li, M.; Bin, R. Exploring AIS data for intelligent maritime routes extraction. *Appl. Ocean. Res.* **2020**, *101*, 102271. [[CrossRef](#)]
6. Lei, P.-R. Mining maritime traffic conflict trajectories from a massive AIS data. *Knowl. Inf. Syst.* **2019**, *62*, 259–285. [[CrossRef](#)]
7. Li, H.; Liu, J.; Liu, R.W.; Xiong, N.; Wu, K.; Kim, T.-H. A Dimensionality Reduction-Based Multi-Step Clustering Method for Robust Vessel Trajectory Analysis. *Sensors* **2017**, *17*, 1792. [[CrossRef](#)]
8. Zhang, T.; Ramakrishnan, R.; Livny, M. BIRCH: An efficient data clustering method for very large databases. *ACM Sigmod Rec.* **1996**, *25*, 103–114. [[CrossRef](#)]
9. Gulati, H.; Singh, P. Clustering techniques in data mining: A comparison. In Proceedings of the 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 11–13 March 2015.
10. Wu, L.; Xu, Y.; Wang, Q.; Wang, F.; Xu, Z. Mapping Global Shipping Density from AIS Data. *J. Navig.* **2016**, *70*, 67–81. [[CrossRef](#)]
11. Yan, Z.; Cheng, L.; He, R.; Yang, H. Extracting ship stopping information from AIS data. *Ocean Eng.* **2022**, *250*, 111004. [[CrossRef](#)]
12. Rong, H.; Teixeira, A.; Soares, C.G. Spatial correlation analysis of near ship collision hotspots with local maritime traffic characteristics. *Reliab. Eng. Syst. Saf.* **2021**, *209*, 107463. [[CrossRef](#)]
13. Yan, Z.; He, R.; Ruan, X.; Yang, H. Footprints of fishing vessels in Chinese waters based on automatic identification system data. *J. Sea Res.* **2022**, *187*, 102255. [[CrossRef](#)]
14. Yan, Z.; Xiao, Y.; Cheng, L.; Chen, S.; Zhou, X.; Ruan, X.; Li, M.; He, R.; Ran, B. Analysis of global marine oil trade based on automatic identification system (AIS) data. *J. Transp. Geogr.* **2020**, *83*, 102637. [[CrossRef](#)]
15. Andersson, P.; Ivehammar, P. Dynamic route planning in the Baltic Sea Region—A cost-benefit analysis based on AIS data. *Marit. Econ. Logist.* **2017**, *19*, 631–649. [[CrossRef](#)]
16. Lei, P.-R.; Tsai, T.-H.; Peng, W.-C. Discovering maritime traffic route from AIS network. In Proceedings of the 2016 18th Asia-Pacific Network Operations and Management Symposium (APNOMS), Kanazawa, Japan, 5–7 October 2016.
17. Altan, Y.C.; Otay, E.N. Maritime Traffic Analysis of the Strait of Istanbul based on AIS data. *J. Navig.* **2017**, *70*, 1367–1382. [[CrossRef](#)]

18. Yu, Q.; Liu, K.; Teixeira, A.; Soares, C.G. Assessment of the influence of offshore wind farms on ship traffic flow based on AIS data. *J. Navig.* **2020**, *73*, 131–148. [[CrossRef](#)]
19. Wei, Z.; Xie, X.; Zhang, X. AIS trajectory simplification algorithm considering ship behaviours. *Ocean Eng.* **2020**, *216*, 108086. [[CrossRef](#)]
20. Zhen, R.; Jin, Y.; Hu, Q.; Shao, Z.; Nikitakos, N. Maritime Anomaly Detection within Coastal Waters Based on Vessel Trajectory Clustering and Naive Bayes Classifier. *J. Navig.* **2017**, *70*, 648–670. [[CrossRef](#)]
21. Wang, C.; Li, G.; Han, P.; Osen, O.; Zhang, H. Impacts of COVID-19 on Ship Behaviours in Port Area: An AIS Data-Based Pattern Recognition Approach. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 25127–25138. [[CrossRef](#)]
22. Zhou, Y.; Daamen, W.; Vellinga, T.; Hoogendoorn, S.P. Ship classification based on ship behavior clustering from AIS data. *Ocean Eng.* **2019**, *175*, 176–187. [[CrossRef](#)]
23. Rong, H.; Teixeira, A.; Soares, C.G. Data mining approach to shipping route characterization and anomaly detection based on AIS data. *Ocean Eng.* **2020**, *198*, 106936. [[CrossRef](#)]
24. Yitao, W.; Lei, Y.; Xin, S. Route mining from satellite-AIS data using density-based clustering algorithm. *J. Phys. Conf. Ser.* **2020**, *1616*, 012017. [[CrossRef](#)]
25. He, Y.K.; Zhang, D.; Zhang, J.F.; Li, T.W. Ship route planning using historical trajectories derived from AIS data. *Trans. Nav. Int. J. Mar. Navig. Saf. Sea Transp.* **2019**, *13*, 69–76. [[CrossRef](#)]
26. Zhao, L.; Shi, G.; Yang, J. Ship trajectories pre-processing based on AIS data. *J. Navig.* **2018**, *71*, 1210–1230. [[CrossRef](#)]
27. Jurdana, I.; Lopac, N.; Wakabayashi, N.; Liu, H. Shipboard Data Compression Method for Sustainable Real-Time Maritime Communication in Remote Voyage Monitoring of Autonomous Ships. *Sustainability* **2021**, *13*, 8264. [[CrossRef](#)]
28. Kurekin, A.A.; Loveday, B.R.; Clements, O.; Quartly, G.D.; Miller, P.I.; Wiafe, G.; Agyekum, K.A. Operational Monitoring of Illegal Fishing in Ghana through Exploitation of Satellite Earth Observation and AIS Data. *Remote Sens.* **2019**, *11*, 293. [[CrossRef](#)]
29. Liu, H.; Jurdana, I.; Lopac, N.; Wakabayashi, N. BlueNavi: A Microservices Architecture-Styled Platform Providing Maritime Information. *Sustainability* **2022**, *14*, 2173. [[CrossRef](#)]
30. Xiao, G.; Wang, T.; Chen, X.; Zhou, L. Evaluation of Ship Pollutant Emissions in the Ports of Los Angeles and Long Beach. *J. Mar. Sci. Eng.* **2022**, *10*, 1206. [[CrossRef](#)]
31. Suo, Y.; Chen, W.; Claramunt, C.; Yang, S. A ship trajectory prediction framework based on a recurrent neural network. *Sensors* **2020**, *20*, 5133. [[CrossRef](#)]
32. Chen, P.; Li, M.; Mou, J. A velocity obstacle-based real-time regional ship collision risk analysis method. *J. Mar. Sci. Eng.* **2021**, *9*, 428. [[CrossRef](#)]
33. Wan, Z.; Ji, S.; Liu, Y.; Zhang, Q.; Chen, J.; Wang, Q. Shipping emission inventories in China's Bohai Bay, Yangtze River Delta, and Pearl River Delta in 2018. *Mar. Pollut. Bull.* **2020**, *151*, 110882. [[CrossRef](#)]
34. NGA. *World Port Index*; National Geospatial-Intelligence Agency: Springfield, IL, USA, 2019.
35. Wang, L.; Chen, P.; Chen, L.; Mou, J. Ship AIS Trajectory Clustering: An HDBSCAN-Based Approach. *J. Mar. Sci. Eng.* **2021**, *9*, 566. [[CrossRef](#)]
36. Dinh, D.T.; Fujinami, T.; Huynh, V.N. Estimating the optimal number of clusters in categorical data clustering by silhouette coefficient. In *International Symposium on Knowledge and Systems Sciences*; Springer: Berlin/Heidelberg, Germany, 2019.
37. Xiao, J.; Lu, J.; Li, X. Davies Bouldin Index based hierarchical initialization K-means. *Intell. Data Anal.* **2017**, *21*, 1327–1338. [[CrossRef](#)]
38. Lorbeer, B.; Kosareva, A.; Deva, B.; Softić, D.; Ruppel, P.; Küpper, A. Variations on the clustering algorithm BIRCH. *Big Data Res.* **2018**, *11*, 44–53. [[CrossRef](#)]
39. Admiralty. *Ocean Passages for the World: Np136*; United Kingdom Hydrograph Office: Taunton, UK, 2018.
40. Likas, A.; Vlassis, N.; Verbeek, J.J. The global k-means clustering algorithm. *Pattern Recognit.* **2003**, *36*, 451–461. [[CrossRef](#)]
41. Birant, D.; Kut, A. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data Knowl. Eng.* **2007**, *60*, 208–221. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.