Journal of
*Marine Science and Engineering*

MDPI

*Article*

# Spatiotemporal Companion Pattern (STCP) Mining of Ships Based on Trajectory Features

**Chunhui Zhou [1,2,3], Guangya Liu [1], Liang Huang [2,4,*] and Yuanqiao Wen [2,3,4]**

1 School of Navigation, Wuhan University of Technology, Wuhan 430063, China
2 National Engineering Research Center for Water Transport Safety, Wuhan 430063, China
3 Hubei Key Laboratory of Inland Shipping Technology, Wuhan 430063, China
4 Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063, China
* Correspondence: leung.huang@whut.edu.cn; Tel.: +86-13477070410

**Abstract:** Spatiotemporal companion pattern (STCP) mining is one of the means to identify and detect group behavioral activities. To detect the spatiotemporal traveling pattern of ships from massive spatiotemporal trajectory data and to understand the movement law of group ships, this article proposes a feature-driven approach for STCP mining that consists of (1) generating the grid index via the rasterizing of geospace and characterizing trajectory points via the spatiotemporal trajectory grid sequences (STTGSs) of ships; (2) designing filtering rules with the constraints of range, time and distance to construct a candidate set for ship STCP mining; and (3) measuring the STTGS similarity of the associated ships and setting the confidence threshold to realize spatiotemporal companion mining. The effectiveness of the proposed method is practically validated on a real trajectory dataset which is collected from the Taiwan Strait waters. The experimental results are as follows: 825 pairs of associated ships and 225 pairs of accompanying ships are mined when the grid size is 0.05° and the confidence is 0.5. Larger grid sizes can increase the inclusiveness of the associated ship trajectory similarity measurement, which can result in an increase in confidence of pattern. A large number of pseudo-accompaniment ships are extracted to the result set, resulting in a more dispersed distribution of pattern confidence. By verifying the proposed method, accompanying behavioral activities such as ship cooperative operation, companion navigation method, and so on, can be detected. These results can provide a reference for the research of ship group behavior identification and have an important application value for water transportation management.

**Keywords:** spatiotemporal data mining; ship companion pattern; multi-feature grid sequences; LCSs (longest common subsequence); AIS data



**Citation:** Zhou, C.; Liu, G.; Huang, L.; Wen, Y. Spatiotemporal Companion Pattern (STCP) Mining of Ships Based on Trajectory Features. *J. Mar. Sci. Eng.* **2023**, *11*, 528. https://doi.org/10.3390/jmse11030528

Academic Editor: Carlos Guedes Soares

Received: 3 December 2022
Revised: 17 January 2023
Accepted: 24 January 2023
Published: 28 February 2023



## 1. Introduction

The spatiotemporal trajectory is a curve in the spatial location dimension and time dimension. The spatiotemporal trajectory data contain rich spatiotemporal features of moving targets, activity characteristics and other valuable information [1–3], which provides a research basis for analyzing the features of moving targets' group activities and mining group behavior patterns. For example, in terms of ship behavior supervision, mining the STCP of ships suspected of crimes can assist the relevant departments of Customs to arrest smugglers on the water. Mining the STCP of engineering vessels such as sand dredges can help identify coordinated operations on the water to optimize traffic organization or stop illegal operations in time. In terms of a maritime safety strategy, mining the STCP of suspicious naval ships facilitates the ability of the national security department to analyze the strategic intent of enemy activities and the current situation, make timely responses, etc.

Automatic identification system (AIS) data have proven to be a valuable source of maritime situational awareness and ship behavior analysis using big data mining. Rich achievements have been made in several AIS data mining-based research directions. To identify the ship activity intention and achieve water traffic situational awareness, some scholars have researched ship trajectory prediction [4–6]. In addition, some scholars suggested a ship behavioral features mining method [7–9] by using machine learning or deep learning to cluster and classify AIS data, which would be helpful to master the rules and characteristics of ship activities. To monitor and supervise the behavioral activities of ships, some scholars have proposed methods of ship collision prevention [10], ship abnormal behavior detection [11] and ship emission inventory calculation [12], which are of great significance for navigation safety and environmental protection on water. More importantly, AIS data can be used to evaluate the operational risk management strategies of shipping companies [13] in the shipping industry, which provides significant enlightenment for shipping risk management, etc. In recent years, the rapid growth of spatiotemporal trajectory data collected by the shipborne AIS has brought both opportunities and challenges to STCP mining. Through mining associated knowledge from spatiotemporal trajectory data [14,15], analyzing features of ship group activities and mining potential group motion patterns, group activity laws and the movement trends of ships can be discovered to provide effective analysis and calculation approaches for ship behavior prediction and abnormality detection [16,17], and help predict maritime group events. These are of great significance for the improvement of maritime traffic safety, optimization of traffic organization and evaluation of water traffic situations.

To analyze ship association relationships accurately and mine the ship STCP quickly from massive AIS data, this paper proposes a spatiotemporal feature-driven approach for ship STCP detection. The other parts of this paper are organized as follows. Section 2 analyzes and concludes the current status of domestic and international research on STCP mining. Section 3 mainly defines the ship STCP by analyzing the ship activity features. Section 4 shows the general process of generating the sample trajectory data by dataset preprocessing, then introduces how to utilize time, space and range as constraints to design the filtering rules and describes how to measure the similarity of associated ship STTGSs to mine the ship STCP. Figure 1 illustrates the framework of the entire processing. Section 5 conducts the algorithm validation experiment and sensitivity analysis experiment using AIS data. Finally, the conclusions and evaluation are drawn in Section 6.

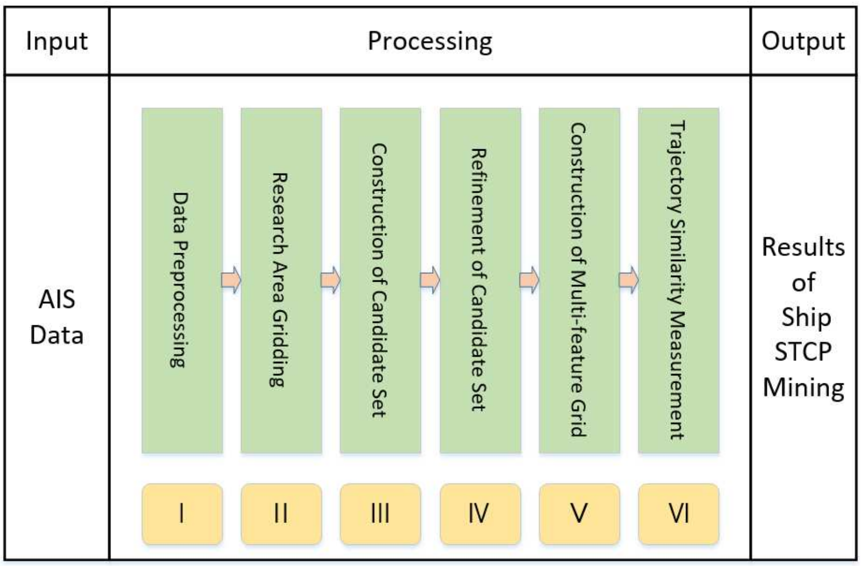



**Figure 1.** The framework of the entire processing.

## 2. Related Work

The STCP of group targets' movements is manifested as groups of spatial objects moving together for a certain period of time [18,19]. This pattern abounds in water traffic scenarios, such as fishing vessels teaming up for fishing operations, offshore engineering vessel collaborative working, military fleets escorting merchant ships, naval vessel formation navigation, etc. To study STCPs of moving objects, researchers have proposed a series of pattern models as well as mining methods. The subjects include vehicles, pedestrians and ships. The existing research results [20–25] are mainly obtained by prolongating the spatial co-location pattern in the temporal dimension. And the mining methods are mainly divided into association rules-based [26] and clustering analysis-based [21].

Not very much work, however, has been devoted to waterborne traffic and major efforts have been directed at vehicles and pedestrians on the road. In the research of road traffic, due to the constraint of the road network, the vehicle travel route has obvious boundaries and the accompanying vehicle detection methods are mainly divided into those based on vehicle GPS trajectory data and those based on ANPR data. The literature [20,21,23,27,28] performed a clustering and correlation analysis on vehicle GPS track data to detect accompanying vehicles. Thi Thi Shein et al. [23] proposed a micro-group clustering method to reduce the time complexity of the clustering algorithm; Zhang Yongmei et al. [27] proposed a two-layer network-based spatiotemporal co-occurrence pattern mining algorithm. It constructs a two-layer spatiotemporal network to store the spatiotemporal proximity relationships between instances and then extracts the proximity relationships in the network to mine the spatiotemporal co-occurrence patterns of vehicles. Concerning the problem that some vehicles do not install or turn off the GPS, the literature [29–32] proposed a vehicle travel companion detection method based on ANPR data, which obtains the license plate and spatiotemporal information of vehicles through traffic cameras and measures the similarity of driving road sections to achieve the mining of accompanying vehicles. Abdulrahman Al-badwi et al. [29] designed a hybrid distributed breadth-first and depth-first frequent itemset mining algorithm HD-FIM based on the Spark platform to reduce the algorithm time complexity from the perspective of improving the Apriori algorithm. Zhu Meiling et al. [30,32] proposed COINCIDENT, an accompanying vehicle detection method based on ANPR data, to achieve the real-time detection of accompanying vehicles. In water traffic, there is no specific roadway constraint for ship navigation and thus the accompanying vehicle detection method does not apply to ship companion pattern mining.

In pedestrian concomitant detection, the mainstream mining method is to time-slice the pedestrian location data and then mine pedestrian companions by the clustering-taking intersection or trajectory similarity metrics. In 2019, Yao Ruihong et al. [33,34] proposed a method of using density clustering combined with association analysis to achieve travel companion discovery for the pattern omission problem caused by sparse trajectories. In 2020, they proposed a travel companion detection framework GroupSeeker for the pattern candidate omission problem caused by short time-slicing and achieved companion detection by density clustering and pseudo-accompaniment filtering. Yu Jiangang et al. [35] designed a Spatiotemporal Trajectory Companion Detection Framework (STCDF) by measuring the trajectory similarities to detect travel companions in trajectories with different sampling frequencies, demonstrating the improvement of the pattern detection efficiency of grid processing. Elahe Naserian et al. [36,37] proposed a loose travel companion pattern that reduced the spatiotemporal continuity of companions and subsequently proposed various clustering-centered detection methods. As the pedestrian trajectory data are obtained through cell phone positioning base stations where the sampling area is relatively fixed, only the spatiotemporal information of pedestrians is retained in the data, while the information on motion features is less so. Therefore, the association strength of travelling companions mined through the clustering-taking intersection method is low. A large number of ship motion features are retained in the ship trajectory, but the variety of ship activities is relatively simple and methods such as clustering cannot meet the requirements for detecting strong STCP.

In the research of water transportation, marine navigation has larger ship sizes, simpler navigation routes and no road network constraints compared with road traffic. The mining method mainly adopts the idea of Apriori association rules to detect the spatiotemporal patterns of ships by mining the spatiotemporal proximity between ships. For large-scale ship trajectory data, Bao Lei [26] proposed a ship spatiotemporal co-occurrence pattern algorithm designed by connecting the instances satisfying spatiotemporal proximity and setting the support and confidence thresholds. He also verified them using the actual AIS dataset on the spatial Hadoop analysis platform architecture. Wang Jiang et al. [38] constructed spatiotemporal matrices and spatiotemporal tensors from motion features and imposed non-negativity constraints and sparsity constraints using a collaborative clustering model to mine the spatiotemporal co-occurrence patterns of ships and reveal the association of ships in the region, time and of ship types. Zhang Yalun et al. [39] proposed a ship spatiotemporal co-occurrence pattern mining method (SSC-IS) based on sliding space–time rectangular and improved support for the pattern omission and computational inefficiency problems in traditional pattern mining methods. These methods either traverse the temporal and spatial data for a large number of point-pair calculations or directly process the spatiotemporal data with high time complexity, which cannot realize the mining of the accompanying patterns of ships, so it is necessary to study suitable methods for the accompanying pattern mining of ship trajectories.

These pattern models are microscopically designed to identify objects that move together for some time. However, to be applied to different scenarios, many kinds of STCPs are proposed, such as Flock, Convoy, Swarm, Platoon and so on. Their concepts have small differences in the continuity of the time period or moving together. The relevance of objects in the STCP pattern mining method based on association rules is stronger than that in the clustering method, but the method has a high requirement on the sampling frequency of the moving target position information. In water traffic, it is difficult to achieve complete synchronization of the trajectory acquisition due to equipment differences and signal strength. The mining algorithm based on the cluster analysis has strong robustness to heterogeneous trajectory data, but it needs to cluster the moving clusters on each timestamp, which has large time overhead and low mining efficiency. The efficiency and accuracy of these methods have a great challenge when faced with massive spatiotemporal trajectory data [40].

## 3. Correlation Analysis and Basic Definitions

Compared with other moving targets (vehicles, pedestrians, animals, etc.), the ship has the characteristics of a large size, large navigational safety space, stable spatial distribution of trajectory data, simple navigational routes, etc. Therefore, when defining the ship STCP, the setting of the proximity distance and the proximity time is different from the existing STCP. This paper generates spatial grid indexes by meshing AIS data. The proximity relationship for distance is considered to be satisfied when ships are in the same grid, the proximity relationship for time is considered to be satisfied when the common time in the same grid meets the duration threshold and the range constraint is considered to be satisfied when the number of common navigational grids meets the threshold.

**Definition 1 (Ship Trajectory Data):** *The raw AIS dataset can be expressed as* $TR = \{Tr_1, Tr_2, Tr_3 \ldots Tr_m\}$*, where m represents the number of ships. Single ship trajectory can be expressed as* $Tr = \{tr_1, tr_2, tr_3 \ldots tr_n\}$*, where n represents the number of trajectory sampling points. A spatiotemporal position point is the vector of the ship instantaneous motion state, which can be expressed as* $tr = \{mmsi, time, lat, lon, cog, heading, sog\}$*, where mmsi represents the Maritime Mobile Service Identify; time represents the timestamp of the trajectory sampling point; lat and lon represent the latitude and longitude of sampling location of the ship trajectory point; cog, heading and sog represent, respectively, the course over ground, the bow direction and the speed over ground of the ship trajectory point.*

**Definition 2 (Consecutively Repeated Elements of Sequence De-duplication):** *All elements in the sequence are de-duplicated by a sliding window of size 2. Only one element is retained when the elements in the sliding window are the same. By sequence de-duplication, the length of* *the sequence can be shortened while the original order of the elements is preserved. For example, given a sequence as* $Seq = \{a, b, c, c, c, d, e, e, c, b, b\}$ *after consecutive element de-duplication of the sequence, a new sequence is obtained as* $Seq_{new} = \{a,\ b,\ c,\ d,\ e,\ c,\ b\}$*.*

**Definition 3 (Support):** *The support of the ship STCP is the number of common elements of the associated ship's STTGSs. Given any two STTGSs as* $GSs_{Tr_A}$ *and* $GSs_{Tr_B}$*, it can be expressed as*

$$\text{support} = GSs_{Tr_A} \cap GSs_{Tr_B} \tag{1}$$

**Definition 4 (Confidence):** *The confidence of the ship STCP is the proportion of the number of common elements of the associated ships' STTGSs to the length of the full cycle associated trajectory sequence, namely the associated vessel concomitant participation rate. It can be expressed as*

$$\text{Confidence} = \text{support}/Tr_{associated} \tag{2}$$

**Definition 5 (Ship STCP):** *The ship STCP means the associated ships have sailed* $\Delta L$ *common multi-feature grids within a certain time* $\Delta T$ *and the proportion of the number of common grids in the associated trajectory sequence exceeds the constraint of the confidence threshold* $\varepsilon$*. Given any pair of associated ship trajectories as* $Tr_A$ *and* $Tr_B$*, respectively, the spatiotemporal features of the ship STCP can be expressed as:*

$$\begin{gathered} Time_{Tr_A} \cap Time_{Tr_B} \geq \Delta T \\ \text{Support} \geq \Delta L \\ \text{Confidence} \geq \varepsilon \end{gathered} \tag{3}$$

## 4. Methods

The STCP mining is the process of extracting accompanying objects from spatiotemporal trajectory data. This paper designs a STCP mining method based on spatiotemporal trajectory features, which is divided into the following steps: firstly, preprocess the AIS data, generate a sample trajectory for STCP mining, set the minimum speed threshold to extract ship navigation trajectories and grid the trajectory dataset to generate spatiotemporal grid indexes. Then, design the filtering rules with the constraints of range, time and distance and generate the more strongly associated candidate set. Finally, the ship STCP mining is realized by measuring the multi-feature grid sequence trajectory similarity of associated ships. The basic framework is shown in Figure 2.

### *4.1. Grid Index Generation*

Due to the interference of positioning equipment signal fluctuations and other uncontrollable factors, there are many outlier points, redundant points and useless data in the trajectory data. Trajectory preprocessing is an indispensable step to achieve efficient spatiotemporal pattern mining [2]. By preprocessing the raw AIS data, the worthless trajectory points can be removed while the raw ship motion information is retained, which reduces the data volume and improves the mining efficiency. The grid index is generated by meshing ship trajectory data and STTGS is used to characterize position point data, which is convenient for the trajectory data to be counted and analyzed later based on the grid index. The specific steps are as follows.

(1) Spatiotemporal Trajectory Preprocessing

The raw AIS data are cleaned to eliminate noise data, error data and redundant data in the trajectory dataset, where the noise data mainly refer to the drifting trajectory points generated by the AIS positioning error, the error data mainly refer to the blatantly false data that do not meet the requirements of each field and the redundant data mainly include static information data and duplicated data that are not related to pattern mining. The sample trajectories of the ship STCP mining are generated after preprocessing.

(2) Navigational Trajectory Extraction

By setting the constraint of a minimum speed, the stop points and anchoring trajectory points in the data are eliminated and the trajectory data of the ship sailing status are generated. The sailing trajectories are grouped by ship's MMSI and serialized in the order of trajectory point sampling time. Sparse trajectory segments are processed with linear interpolation to improve the integrity of the track data.

(3) Spatial Grid Index Construction

The global geographic location is rasterized and the research area is divided into uniform square grids with starting latitude and longitude of $-180°$ and $-90°$, respectively. The cleaned spatiotemporal trajectory dataset is mapped into the grids to generate spatiotemporal grid indexes and obtain the STTGSs. For example, given a grid size of $\theta$ and the raw trajectory $tr = \{mmsi, time, lat, lon, cog, heading, sog\}$, the generated grid index is defined as follows:

$$x = [(lon - \left(-180 - \frac{1}{2} * \theta\right))/\theta] \tag{4}$$

$$y = [(lat - \left(-90 - \frac{1}{2} * \theta\right))/\theta] \tag{5}$$

After rectification, the equations are shown respectively in Equations (6) and (7):

$$x = [(lon + 180)/\theta + \frac{1}{2}] \tag{6}$$

$$y = [(lat + 90)/\theta + \frac{1}{2}] \tag{7}$$

$x$ and $y$ represent the indexes of the grid where the latitude and longitude of the trajectory points are located. Both start counting from "1" and [] represents rounding the value down, which is called the integer part of the number. By generating the grid indexes, the raw trajectory is characterized by the spatiotemporal grid trajectory sequences, which can be expressed as $trgs = \{grid\ id, mmsi, time, lat, lon, cog, heading, sog\}$, and the grid index can be expressed as Grid id $= (x, y)$.

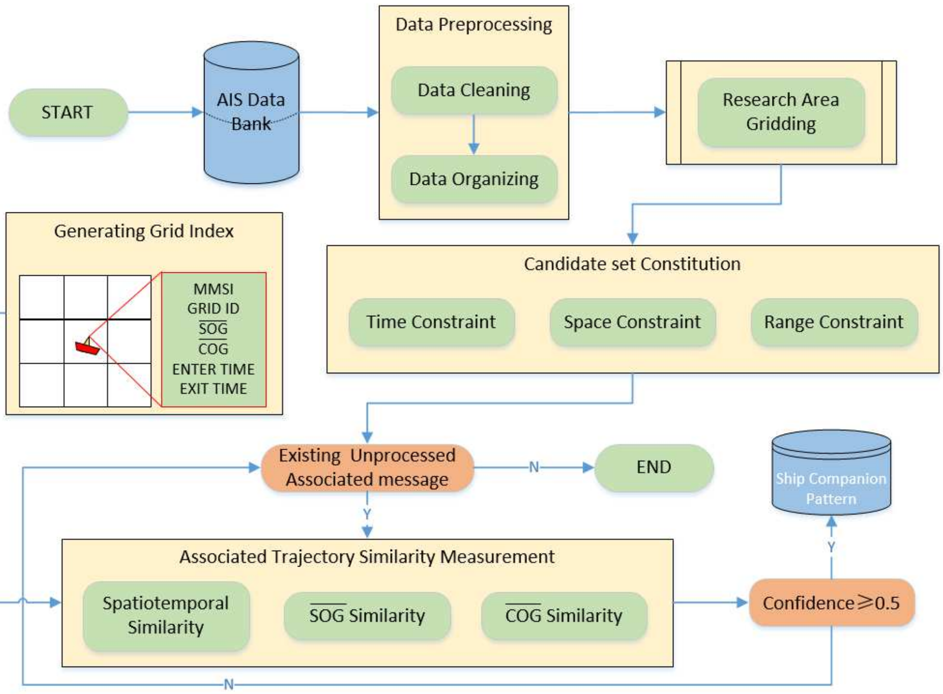



**Figure 2.** The detailed process of the methodology.

### *4.2. Candidate Set Construction*

Although the redundant trajectory points can be basically eliminated with the above process, the scale of the trajectory data to be processed is still large. If pairwise matching is made directly based on spatiotemporal features of ships extracted arbitrarily on MMSI from the AIS data, the inefficient algorithm will not meet the demand for fast mining of massive maritime trajectory data. Therefore, STCP mining is realized by constructing a candidate set of associated ships. The moving targets of STCP navigate range over $\Delta L$ grids together in a certain time $\Delta T$ with a proximity distance $\Delta D$ constraint. The ships in the same grid are regarded as meeting the proximity distance constraint. The filtering rules are designed with time, distance and range as constraints; the associated ships that may have accompanying relationships are extracted and the pattern candidate set is generated. The specific steps are as follows.

(1) Short Spatiotemporal Trajectory Elimination

The short trajectory is characterized by short navigation time and range, so the filter is constructed according to time and range. Firstly, the start and end times of each ship trajectory sampling are counted to judge whether the duration time satisfies the constraint of the accompanying time span $\Delta T$. The trajectories with a consecutive time span less than $\Delta T$ are eliminated and the dataset is updated. Then, the range of ships is filtered. Since the positioning frequency of the shipboard AIS is much less than the time used for a ship to navigate a grid, there will be multiple trajectory points in the same grid for the ship and there are many duplicate elements in the ship STTGSs. A new grid index sequence ($GSs_{-}$) is obtained by removing the repeated grid indexes in the grid sequence of each ship. Finally, the ship trajectories of which the STTGS length ($len(GSs_{-})$) satisfies the constraint of the accompanying range threshold ($\Delta L$) are retained. If not, the trajectories are eliminated. Thus, the ship trajectory dataset is updated.

(2) Initial Candidate Set Generation

The STCP mining is the process of filtering each feature correlation of the ships. To achieve fast mining of ship STCP, the associated ships dataset is constructed with time as the constraint and the initial candidate set for pattern mining is generated by merging these datasets. First, one ship is extracted arbitrarily from the AIS dataset as the primary target and others as the secondary target. The common time span of the primary target and the secondary target trajectory is calculated. The span is filtered by the accompanying time threshold $\Delta T$. After an iterative calculation, the primary and secondary targets satisfying the time constraint are used as spatial-associated ships to generate the associated ships dataset, which is represented as $ICS_{O_A} = \{O_{a1}, \ldots, O_{an}\}$. This can be described as:

$$\begin{gathered} \forall O_a \in \{O_{a1}, \ldots, O_{an}\}, \\ Min(O_A time_{max}, O_a time_{max}) - Max(O_A time_{min}, O_a time_{min}) \geq \Delta T \end{gathered} \tag{8}$$

where $O_A$ and $O_a$ represent the primary ship "$A$" and the secondary ship "$a$", respectively; $O_A time_{max}$ and $O_A time_{min}$ represent the maximum and minimum values of the primary ship A's trajectory sampling time, respectively. Then the whole trajectory dataset is traversed, the associated ships of each ship are calculated iteratively and the condition $|mmsi_{Oa}| > |mmsi_{OA}|$ as the iterative constraint is set to avoid repeating the matching and recording of subsequent ships, where $|mmsi_{OA}|$ represents the numerical value of the MMSI. The associated ships that meet the requirements of the ship STCP about the time characteristic are screened out and merged to generate the initial candidate set for the ship STCP mining. Finally, these messages are stored in fields as $\{mmsi_1, mmsi_2, timestart, timeend\}$, where $timestart$ and $timeend$ represent the start time and end time of the common time span.

(3) Candidate Set Refinement

To strengthen the spatial correlation of associated ships in the initial candidate set, filtering rules are designed with a space feature to generate a strong association candidate set for pattern mining. Firstly, the de-duplicated grid sequences ($GSs_{-}$) of each ship in the initial candidate set are extracted and the number of common elements in their grid index sequences is calculated. Then the accompanying range threshold ($\Delta L$) is employed to refine the initial candidate set. After filtering and iterative calculations, the associated ships satisfying the constraint are saved as pattern mining objects in a new candidate set, which is represented as $CS_{O_A} = \{O_{a1}, \ldots, O_{am}\}$, i.e., $\forall O_a \in \{O_{a1}, \ldots, O_{am}\}$, which both satisfy

$$GSs_{O_A-} \cap GSs_{O_a-} \geq \Delta L \tag{9}$$

where $GSs_{O_A-}$ and $GSs_{O_a-}$ represent the de-duplicated grid index sequences of the primary ship "$A$" and the secondary ship "$a$". Finally, by traversing and updating the initial candidate set, the associated ships with low relevance in terms of the space feature can be eliminated and the strong association candidate set can be obtained by refining the initial candidate set. The field format is the same as the initial candidate set. The multiple filtering rules are used to eliminate the interfering objects of the STCP mining and reduce the data redundancy, thus improving the efficiency and accuracy of mining.

### *4.3. Ship STCP Mining*

The spatiotemporal motion features of the associated ships are highly correlated during accompanying navigation. To measure each feature correlation of the associated ships from the strong correlation candidate set and mine the STCP based on the grid index, the improved longest common subsequences (LCSs) algorithm is used to measure the STTGS similarity of the associated ships. The specific steps are as follows.

(1) Multi-motion Attribute Grid Construction

By setting constraints on the spatiotemporal features of the associated ship trajectory, the spatiotemporal correlation of the associated ships is improved, but the pseudo-accompaniment ships may be misjudged as the STCP such as the pursuit crossing behavior, following behavior, etc. To improve the accuracy of local trajectory similarity measurements and pattern mining, it is necessary to constrain the ship motion attributes in each grid, so the average speed and average course constraints are set to improve the motion attribute correlation of locally correlated trajectories. By iterating through the de-duplicated grid sequences of each ship ($GSs_-$), the average speed and average course over ground of the ship in each grid it passes through are calculated. In addition, the times of entering and exiting each grid are counted for the multi-motion attribute grid construction. The information is collected to generate the motion attribute table of each ship, which facilitates the information call when the trajectory similarity measurement is performed. The table is held in six fields, i.e., $\{MMSI,\ GRIDID,\ EN-TIME,\ EX-TIME, \overline{COG},\ \overline{SOG}\}$, as shown in Table 1.

**Table 1.** Multi-motion attribute grid information.

| *MMSI* | *GRIDID* | *EN-TIME* | *EX-TIME* | $\overline{COG}$ | $\overline{SOG}$ |
|---|---|---|---|---|---|
| 473,487,616 | (453,367) | 1,549,123,204 | 1,549,124,437 | 30 | 7.8 |
| 473,487,616 | (453,368) | 1,549,124,643 | 1,549,125,901 | 350 | 7.7 |
| - - | - - | - - | - - | - - | - - |
| 984,567,372 | (627,264) | 1,551,542,412 | 1,551,543,656 | 57 | 7.9 |
| 984,567,372 | (628,264) | 1,551,543,967 | 1,551,545,327 | 70 | 9.2 |

(2) Trajectory Distance Measurement of Associated Ships

The traditional longest common subsequence algorithm calculates the number of identical elements in two sets to express the similarity between the two sets. For example, given two non-empty sequence sets X and Y, the longest common subsequence is calculated as:

$$LCSs[i][j] = \begin{cases} 0 & i = 0 \cup j = 0 \\ LCSs[i-1][j-1] + 1 & i, j > 0, x_i = y_j \\ max(LCSs[i-1][j], LCSs[i][j-1]) & i, j > 0, x_i \neq y_j \end{cases} \tag{10}$$

where $LCSs[i][j]$ represents the length of the longest common subsequence between the top $i$ elements of the X set and the top $j$ elements of the Y set. The elements in the above two sets are one-dimensional, so it is necessary to improve the original algorithm so that it can measure the similarity of the multi-motion attribute grid sequences of the associated ships. On the basis of measuring the similarity of grid index sequences, the average course difference and the average speed difference of the associated ships in the same grid must satisfy the constraints of threshold $\Delta\theta$ and $\Delta v$. Thus, the improved longest common subsequence algorithm can be obtained. For example, given a pair of associated ship STTGSs, $trgs_A = \{(x,y)_A, mmsi_A, time_A, lat_A, lon_A, direction_A, heading_A, speed_A\}$, $grid\ id_A = (x,y)_A$, $trgs_a = \{(x,y)_a, mmsi_a, time_a, lat_a, lon_a, direction_a, heading_a, speed_a\}$, $grid\ id_a = (x,y)_a$, the longest common subsequence of the pair ship STTGS are described as:

$$SGT = \min(EXTIME_A, EXTIME_a) - max(ENTIME_A, ENTIME_a) \tag{11}$$

$$LCSs[i][j] = \begin{cases} 0 \quad i = 0 \cup j = 0 \\ LCSs[i-1][j-1] + 1\ i, j > 0, \begin{cases} grid\ id_{Ai} = grid\ id_{aj} \\ SGT \geq t \\ \left|\overline{COG}_{Ai} - \overline{COG}_{aj}\right| \leq \Delta\theta \\ \left|\overline{SOG}_{Ai} - \overline{SOG}_{aj}\right| \leq \Delta v \end{cases} \\ max\{LCSs[i-1][j], LCSs[i][j-1]\} \quad i, j > 0, others \end{cases} \tag{12}$$

where $SGT$ represents the common time of the ships in the same grid, $LCSs[i][j]$ represents the length of the longest common subsequences of the $TR_A$ and $TR_a$. The length of the longest common subsequences adds "1" if and only if all judgment conditions are met. Since the elements of the longest common subsequence appear in the same order in both sets, i.e., the longest common subsequence is a vector. It is considered that the accompanying ships and Head-on ships have the same STTGSs. The course constraint is added to the condition, which can exclude Head-on Situation. The STTGSs distance between the associated ships can be obtained by a calculation.

(3) Ship STCP Mining

Firstly, the longest common subsequence of the associated ship STTGSs can be selected as the support of the STCP. The associated ships whose support does not meet the requirement of the accom panying range $\Delta L$ are eliminated. The pseudo-associated ships with the same grid index but different motion characteristics are eliminated. Then the similarity of the associated trajectory is calculated as follows:

$$sim_{(TR_A, TR_a)} = \left(LCSs_{(TR_A, TR_a)}/len_{TR_A} + LCSs_{(TR_A, TR_a)}/len_{TR_a}\right)/2 \tag{13}$$

$$Tr_{associated} = 2 * \frac{len_{TR_A} * len_{TR_a}}{len_{TR_A} + len_{TR_a}} \tag{14}$$

where $LCSs_{(TR_A, TR_a)}$ represents the length of the longest common subsequences of the $TR_A$ and $TR_a$; $len_{TR_A}$ and $len_{TR_a}$ represent the length of the STTGSs of the $TR_A$ and $TR_a$, respectively. As the length of the grid sequence of the associated ship spatiotemporal trajectories differs, the formula for calculating the length of the associated ship integrated sequences is given in Equation (11). In addition, all parameter notations are listed in Table 2. Finally, the identified binary ship STCP spatiotemporal trajectory information is organized and stored and the STCP of multivariate ships can be mined on this basis.

**Table 2.** Description of parameter notations.

| Notation | Description |
|---|---|
| $TR$ | The entire raw dataset |
| $Tr$ | The spatiotemporal trajectory of a ship |
| $tr$ | The single trajectory point with multiple features |
| $\Delta T$ | The duration threshold for the STCP |
| $\Delta L$ | The range threshold for the STCP |
| $\theta$ | The size of a grid |
| $\Delta\theta$ | The $\overline{\text{COG}}$ difference threshold of the associated ships in the same grid |
| $\Delta v$ | The $\overline{\text{SOG}}$ difference threshold of the associated ships in the same grid |
| $\varepsilon$ | The confidence threshold for the STCP |
| $t$ | The common time threshold of the associated ships in the same grid |
| STTGSs | Spatial-Temporal Trajectory Grid Sequences |
| $Tr_{associated}$ | The representative length of associated ship STTGSs |

## 5. Experimental Results and Discussion

In this section, experiments are conducted with real AIS data to verify the effectiveness of the ship STCP mining methodology proposed in this paper and parameter sensitivity analysis of the grid size is performed.

### 5.1. Dataset and Experimental Environment

The experimental waters are collected from Taiwan Strait waters, China, and the experiments are done on a computer with an Intel Core i5 1035G1 CPU, 16G RAM and 512G SSD and a 64-bit Windows 10 operating system as the software environment. All algorithms for the experiments are implemented using Python and the tools used include Pandas, Numpy, Basemap, Matplotlib and other toolkits. The experiments collect the AIS ship trajectory data from February to mid-April 2019, with a total of 12,994,871 data and the data size of 919 MB. The data are provided by the Fujian Maritime Safety Administration.

### 5.2. Mining Algorithm Validation Experiment

Generating the Spatial Grid index. The AIS data are preprocessed as described in Section 4.1, including data cleaning, data serialization, navigation trajectory extraction and grid index generation. In the process of navigation trajectory extraction, the speed threshold is taken as 0.8 knots to eliminate the stopping point of the anchored, berthing and slow-moving ships. Then, the grid size is taken as 0.05°, about 2.7 NM and the ship trajectory data are gridded to generate the spatiotemporal grid index. After preprocessing with the above method, a total of 5,049,433 spatiotemporal navigation trajectory data are obtained, including 14,413 ships.

Constructing the Candidate Set for Ship STCP Mining. According to the method described in Section 4.2, the short trajectory, which includes the short time span of ship trajectories and short grid sequences after de-duplication, is eliminated first. The consecutive time span threshold of a single ship trajectory sampling $\Delta T$ takes 36,000 s, namely 10 h, and the number of grids occupied by every ship trajectory $\Delta L$ takes 20, about 54NM. Then, the common time of associated ships with the common elements of their STTGSs is matched. The associated ships whose trajectory sampling common time is more than 10 h and the number of common grid indexes is more than 20 are extracted. Finally, the associated ship dataset is merged to generate a strong association candidate set for STCP mining and a total of 1,104,202 pairs of associated ships are obtained.

Ship STCP Mining. According to the method described in Section 4.3, the multi-motion attribute grid is constructed by calculating the motion information of each ship in each grid it passes through. A total of 1,434,264 multi-feature grid data are generated. The improved longest common subsequence algorithm is used to measure the distance of the associated ship STTGSs, where the average speed threshold $\Delta v$ and the average course threshold $\Delta\theta$ are taken as 6 Kn and 30°, respectively, and the common time threshold $t$ of associated ships in the same grid is taken as 30. The number of grids with an average speed difference less than 6 Kn, an average course difference less than 30° and a common time of more than 30 s in the same grid should be more than 20. By verifying the STCP mining method with AIS data, 825 pairs of associated ships are mined. When the confidence threshold $\varepsilon$ of the STCP is taken as 0.5, 225 pairs of accompanying ships are obtained and the information of the top 8 pairs of accompanying ships is shown in Table 3 in descending order of confidence.

**Table 3.** STCP mining results.

| Serial Number | MMSI-ship1 | MMSI-ship2 | Start-Time | End-Time | Lcss | Similar |
|---|---|---|---|---|---|---|
| 1 | 321,321,356 | 414,352,580 | 1,551,121,623 | 1,551,270,956 | 102 | 0.854289 |
| 2 | 321,321,356 | 413,474,870 | 1,551,121,286 | 1,551,274,395 | 105 | 0.848153 |
| 3 | 400,068,068 | 412,520,467 | 1,550,639,223 | 1,550,729,624 | 89 | 0.843621 |
| 4 | 413,474,870 | 414,352,580 | 1,551,121,623 | 1,551,270,956 | 105 | 0.840215 |
| 5 | 412,431,217 | 412,520,384 | 1,550,642,853 | 1,550,729,624 | 98 | 0.813628 |
| 6 | 413,366,240 | 413,368,420 | 1,551,180,591 | 1,551,297,622 | 135 | 0.784884 |
| 7 | 413,425,410 | 419,057,483 | 1,549,919,223 | 1,551,096,423 | 102 | 0.782172 |
| 8 | 412,442,134 | 413,832,568 | 1,549,079,759 | 1,551,336,379 | 254 | 0.762873 |

To verify the effectiveness and accuracy of the algorithm, the spatiotemporal trajectories of the accompanying ships are displayed in a visualized form and the spatiotemporal trajectories of the first 8 pairs of accompanying ships in the result set of the STCP mining are plotted respectively, as shown in Figures 3 and 4. These show respectively the 2D spatial trajectory and 3D spatiotemporal diagram of the 8 pairs of accompanying ship trajectories combined with geographic information. The yellow and green colors represent the accompanying ship trajectories and different line widths are set to show the accompanying ship spatiotemporal trajectory. From the figure, it can be seen that there are some regional differences in position information, speed and course of the accompanying ship trajectory, but the spatiotemporal features and motion characteristics of the accompanying ships are highly correlated and the sailing routes are basically the same.

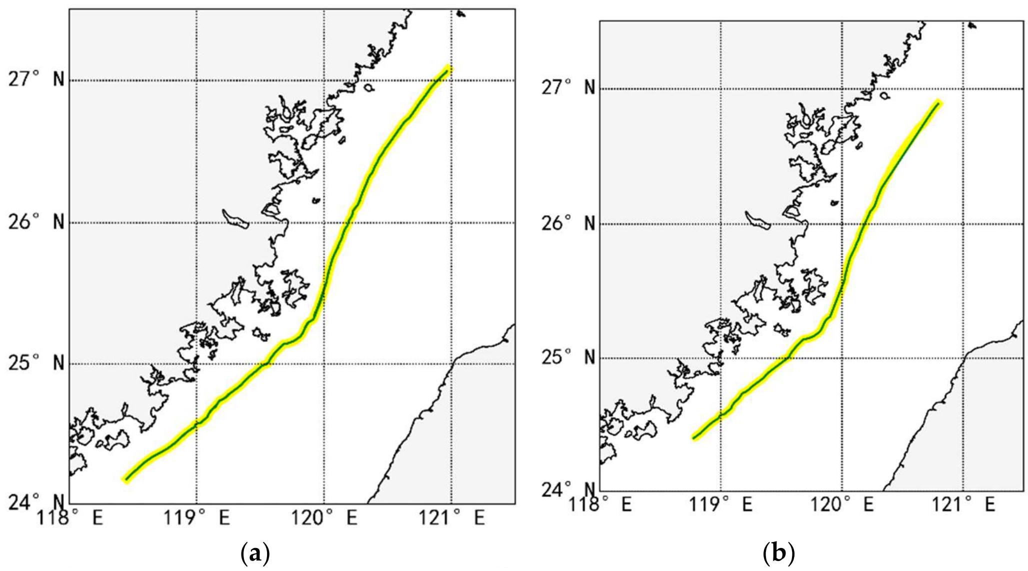



**Figure 3.** *Cont.*

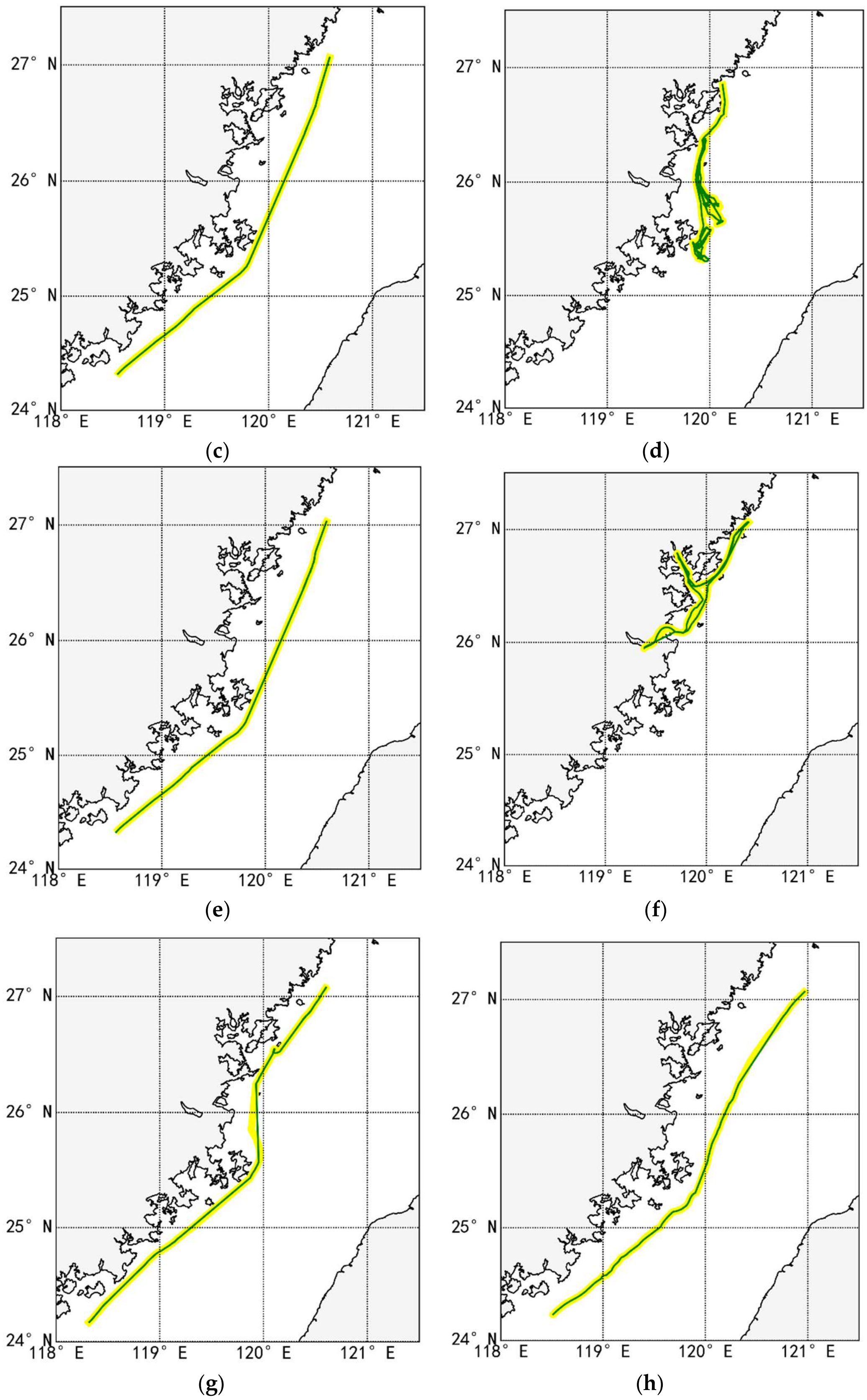



**Figure 3.** STCP trajectory diagram. (**a**–**h**) respectively represent the trajectory of the 8 pairs of accompanying ships in Table 3.

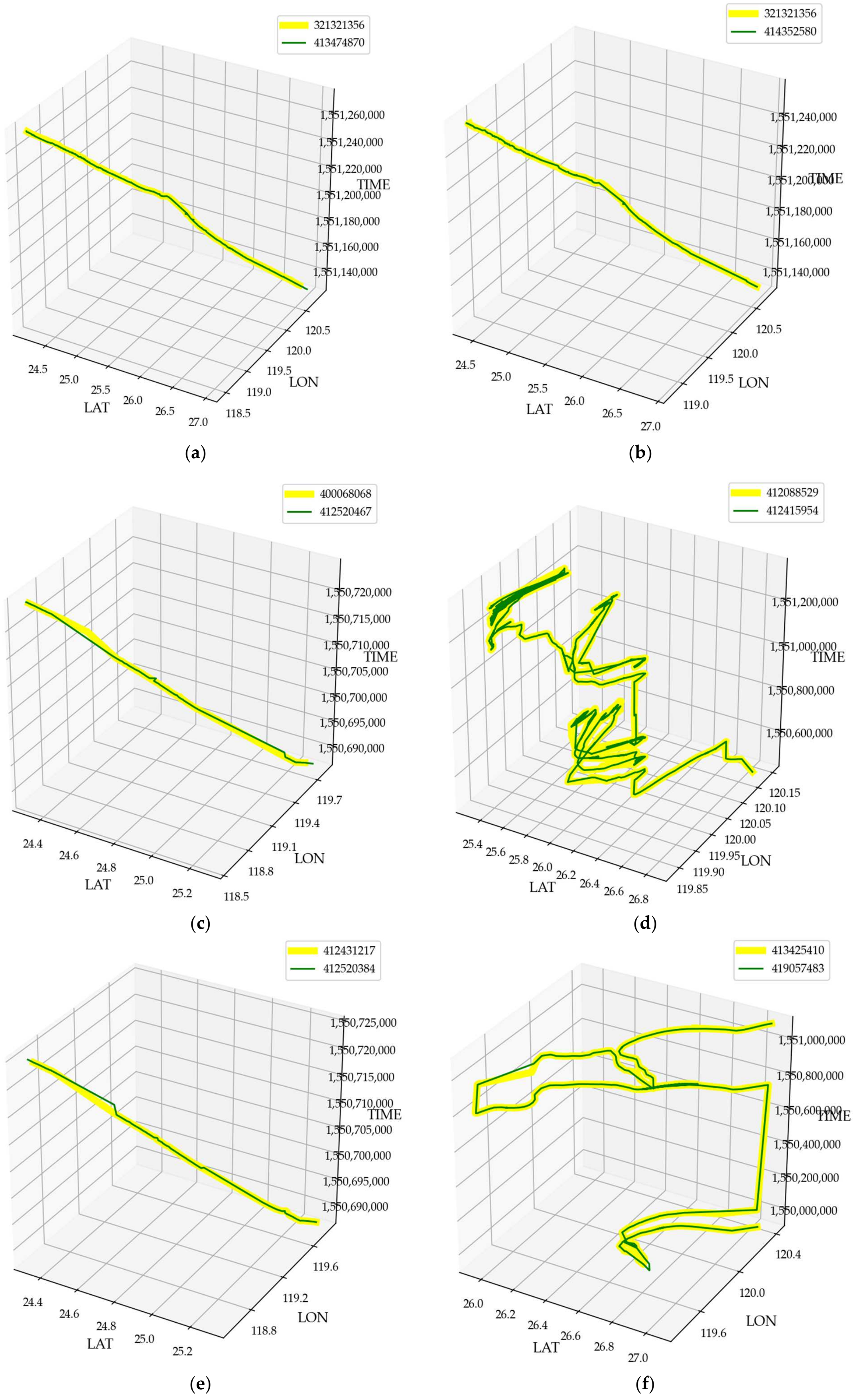
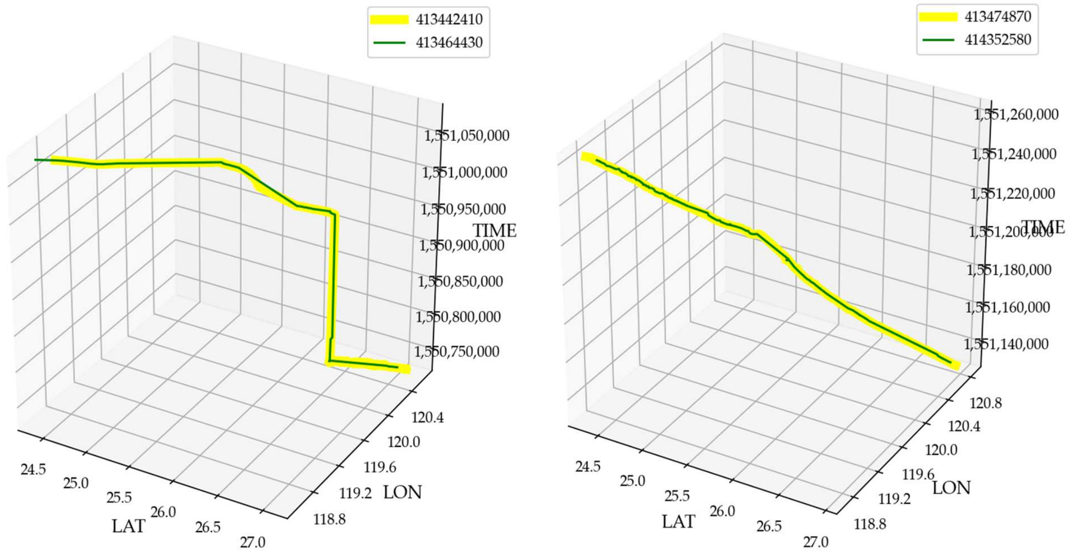



**Figure 4.** *Cont.*




(**g**) (**h**)

**Figure 4.** Spatiotemporal diagram of the STCP. (**a**–**h**) respectively represent the spatiotemporal trajectory of the 8 pairs of associated ships in Table 3.

By visualizing and analyzing the spatiotemporal trajectories of the associated ships, it can be initially verified that the multiple features of the accompanying ships are highly correlated. For further in-depth analysis of the associated ship accompanying behavior, the ship types of the above 8 pairs are searched via the online ship information query website. These are shown in Table 4 and the following inferences can be made with reference to the existing ship types.

(1) The third pair and the fifth pair may be fishing vessels conducting coordinated fishing operations.
(2) The seventh pair may be two cargo vessels navigating together.
(3) The eighth pair may be a fishing vessel and a cargo vessel sailing along the way.

**Table 4.** Type of vessels.

| MMSI | Ship Type | MMSI | Ship Type |
|---|---|---|---|
| 321,321,356 | Other | 414,352,580 | Other |
| 400,068,068 | Fishing | 412,520,467 | Fishing |
| 413,366,240 | Other | 413,368,420 | Cargo |
| 412,431,217 | Fishing | 412,520,384 | Fishing |
| 413,425,410 | Cargo | 412,442,134 | Fishing |
| 413,832,568 | Cargo | 419,057,483 | Cargo |
| 413,474,870 | Tug | | |

Several other pairs of accompanying ships either lacked ship types or did not belong to group activities such as navigation formation or cooperative operation, so they are not considered as key research and analysis subjects.

To analyze the influence of support and confidence on the results of the STCP mining, the relationships between support, confidence and the number of associated ships are shown in Figures 5 and 6, respectively. From Figure 5, it can be seen that with the increase in support, the number of associated ships has a sharp and then slow decrease. From Figure 6, it can be seen that with the increase in confidence, the number of associated ships decreases sharply in the middle part of the confidence. The confidence reflects the reliability of the ship association relationship. The larger the confidence, the more valuable and reliable the association is, which shows the importance of support and confidence in the result of STCP mining.

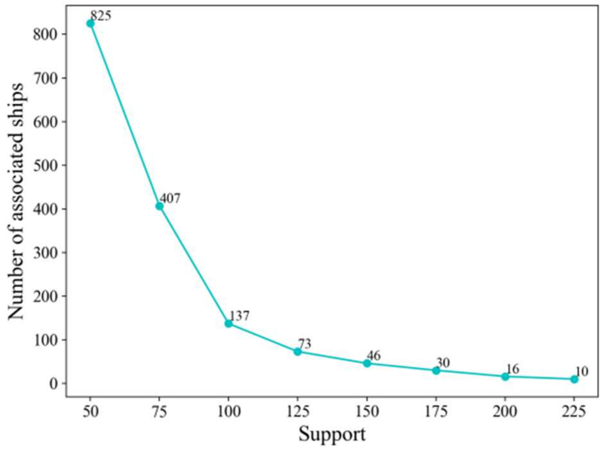



**Figure 5.** Degree of the support impact analysis.

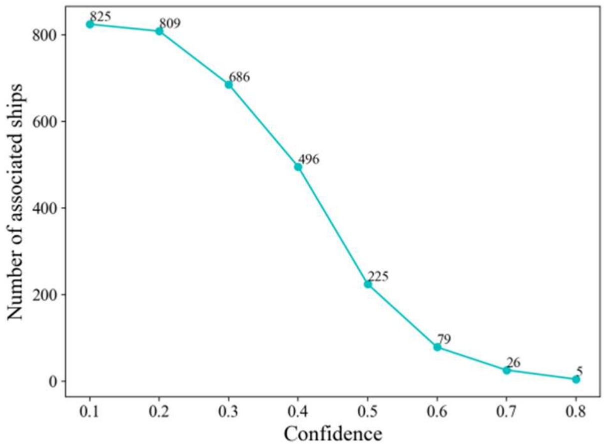



**Figure 6.** Degree of the confidence impact analysis.

The algorithm validation experiments are conducted using real ship trajectory data and the experimental results prove that the proposed methodology can effectively mine the ship STCP with different association strengths. The feasibility and accuracy of the algorithm are demonstrated by analyzing the experimental results in multiple aspects and using visualization methods to show the correlation of multiple features of associated ships, as well as the quantitative relationship between support, confidence and associated ships.

*5.3. Parameter Sensitivity Analysis*

In the application of the STCP mining methodology proposed in this paper, the determination of the grid size is extremely critical. A larger grid size may lead to the mining of pseudo-accompaniment patterns and reduce the accuracy of the algorithm; a smaller grid size results in a longer grid sequence for a single ship, which reduces the mining efficiency and may also lead to a larger difference in the grid index sequences of associated ships and pattern loss. Under the condition that the scale ($\Delta T = 36,000$ s, $\Delta L = 54$ NM) of the ship STCP mining remains unchanged, an important parameter $\Delta L$ of the ship STCP will also change with the change in grid size, which shows the importance of the sensitivity analysis of the grid size. Therefore, this section further probes into the parameter of the grid size and analyzes its effects on the results of STCP mining.

To investigate the effects of different grid sizes on the experimental results, the comparison experiment of grid size is conducted. The dataset size of 919 MB, accompanying duration time threshold $\Delta T = 36,000$ s, the range $\Delta L = 54$ NM and criteria ($\Delta v$, $\Delta\theta$ and $t$) for determining whether the grid is common are kept unchanged, while the grid size (respectively 0.01°, 0.025°, 0.05° and 0.1°) is varied. The other parameters of the comparison experiments and the corresponding experimental results are shown in Table 5.

**Table 5.** Parameter setting and experimental results of the comparison experiment.

| $\theta$ | $\Delta L$ | Candidate Set | Result Set | Pattern Set |
|---|---|---|---|---|
| 0.010° | 100 | 549,535 | 992 | 107 |
| 0.025° | 40 | 2,559,817 | 30,267 | 573 |
| 0.050° | 20 | 5,102,116 | 142,729 | 1556 |
| 0.100° | 10 | 9,156,517 | 600,041 | 8620 |

To analyze the influence of the grid size on the process of STCP mining, the number of associated ships in different grid sizes is counted. The relationship between the number of associated ships and the grid size in the strong association candidate set, the result set and the ship STCP set with confidence greater than 0.5 for STCP mining is shown in Figure 7. It can be seen that with the increase in the grid size, a large number of pseudo-accompaniment patterns that only meet the pattern requirements in terms of spatial and temporal characteristics, but have big differences in terms of motion characteristics are added to the pattern candidate set, which increases the redundancy of the strong association candidate set and thus decreases the efficiency of pattern mining.

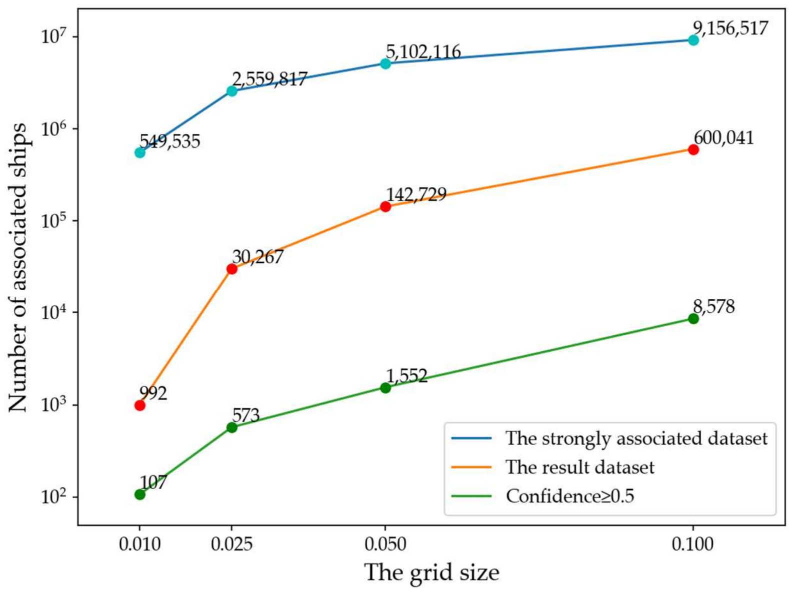
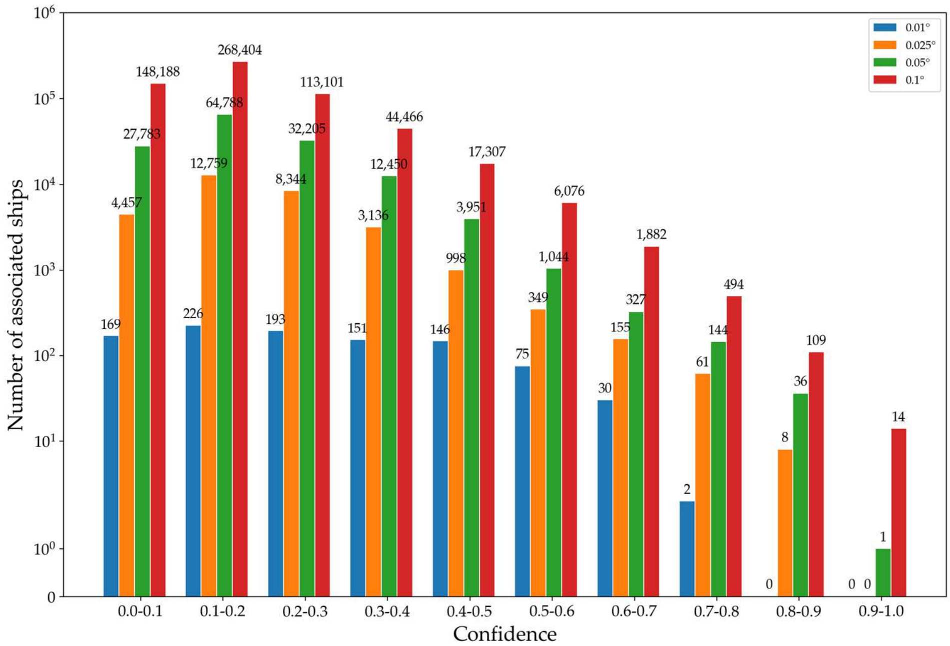



**Figure 7.** Influence analysis of the grid size.

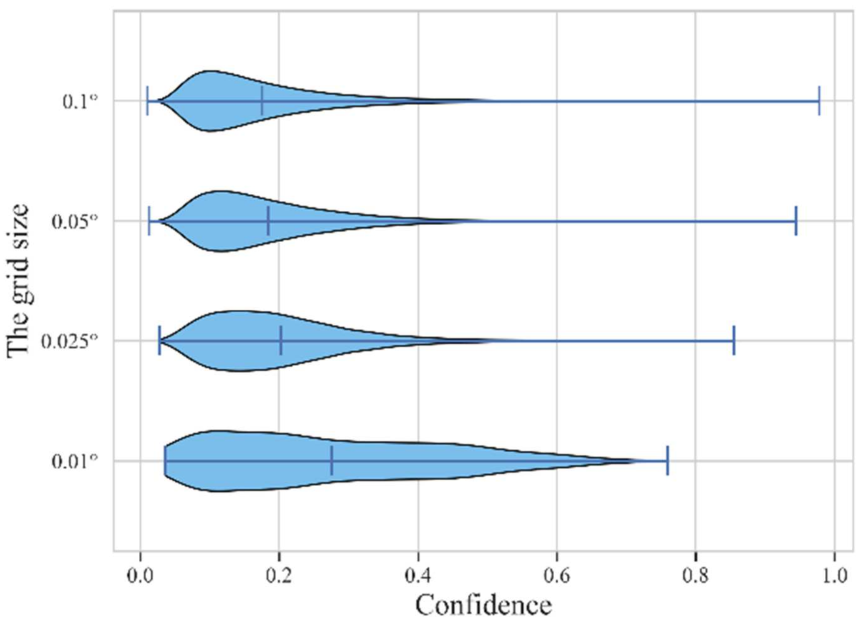
The number of associated ships in every confidence interval in the result has been counted to analyze the effect of the grid size on the confidence, as shown in Figure 8a. The number of associated ships in every confidence interval increases exponentially with the increase in the grid size. The distribution of the confidence data on different grid sizes in the pattern mining result set and the STCP set with the confidence greater than 0.5 is visualized in the form of violin plots, as shown in Figure 8b,c, respectively, where the three lines of violin plots—upper edge, median and lower edge—are from top to bottom. The upper and lower boundaries, upper and lower quartiles and median of Figure 8b,c are shown in Table 6.

**Table 6.** Parameters of Figure 8b,c.

| Figure Number | $\theta$ | Lower Boundary | Lower Quartile | Median | Upper Quartile | Upper Boundary |
|---|---|---|---|---|---|---|
| Figure 8b | 0.01° | 0.036314 | 0.135674 | 0.242330 | 0.402742 | 0.759279 |
| | 0.025° | 0.027306 | 0.124421 | 0.182948 | 0.255838 | 0.854828 |
| | 0.05° | 0.012761 | 0.110687 | 0.161178 | 0.234960 | 0.944444 |
| | 0.1° | 0.010081 | 0.100505 | 0.147445 | 0.221113 | 0.978261 |
| Figure 8c | 0.01° | 0.500958 | 0.532294 | 0.570989 | 0.610566 | 0.759279 |
| | 0.025° | 0.500031 | 0.531680 | 0.571754 | 0.633249 | 0.854828 |
| | 0.05° | 0.5 | 0.523208 | 0.56013 | 0.626344 | 0.944444 |
| | 0.1° | 0.5 | 0.524337 | 0.55774 | 0.610455 | 0.978261 |

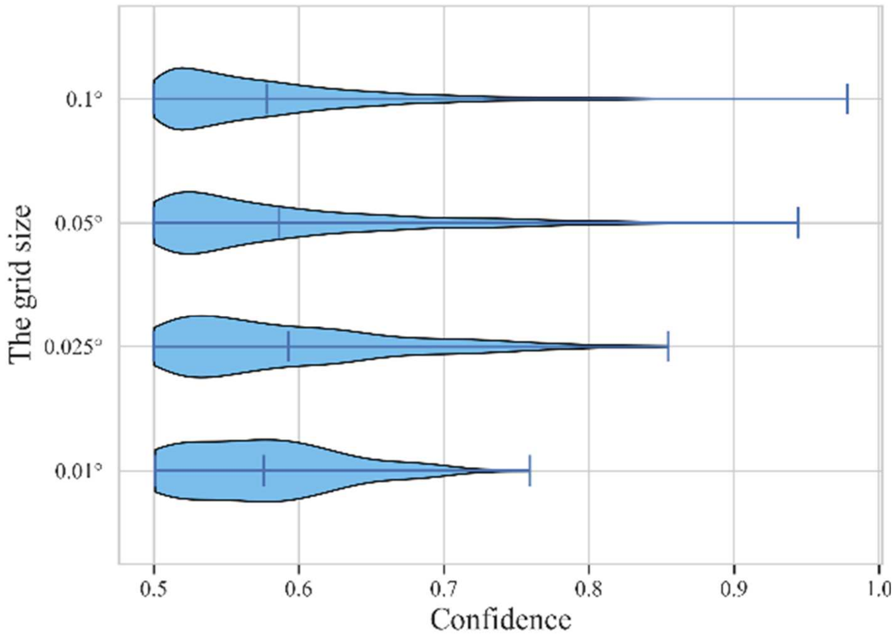



(**a**)




(**b**)

**Figure 8.** *Cont.*




(**c**)

**Figure 8.** The effect of different grid sizes on confidence. (**a**) Quantity distribution of associated ships in each confidence interval of the result set corresponding to each grid size; (**b**) confidence distribution of the result set corresponding to each grid size; (**c**) confidence distribution of the pattern set corresponding to each grid size.

By analyzing Figure 8b,c with the relevant values in Table 6, it can be seen that the confidence in each dataset increases as the grid size increases, implying that increasing the grid size can improve the inclusiveness of the associated ship trajectory similarity measurement, resulting in the increase in the confidence. From Figure 8b, it can be seen that the distribution of the confidence becomes more dispersed as the grid size increases and the median confidence decreases, while the tendency of the overall confidence decreasing is more remarkable. It also reveals that as the grid size increases, a large number of pseudo-accompaniment ships are extracted to the pattern result set, resulting in a decrease in the median of confidence. From Figure 8c, the median of the confidence is the largest when the grid size is 0.025°, which proves that the overall correlation of the ship STCP mining with this grid size is stronger. Through the comprehensive analysis of the experimental results, the grid size of 0.025° is optimal for mining ship STCPs in the experimental area. The above graphical information and analysis show that as the grid size increases, the inclusiveness of the multi-feature grid sequence similarity measurement of the associated ships increases, while the number of low-correlation-associated ships increases, making the confidence of the associated ships in the result set more and more scattered. In addition, the median of the confidence decreases with the increase in the grid size. The number of pseudo-accompaniment ships increases dramatically, which increases the redundancy of the association candidate set and reduces the pattern mining efficiency. Therefore, choosing an appropriate grid size for a specific water area is crucial to accurate and efficient ship STCP mining.

## 6. Conclusions

Ship STCP mining is the process of extracting accompanying ships. The spatiotemporal trajectory feature-driven STCP mining, by setting multiple constraints on trajectory features, calculates the ship association strength of the AIS trajectory data and discovers the patterns and laws behind the association relationship. This is of great research significance for marine traffic safety supervision and special ship group behavior detection. In this paper, a ship STCP-mining approach based on spatiotemporal trajectory features is proposed. The improved longest common subsequence algorithm is used to measure the similarity of the multi-feature grid trajectory sequences of the associated ships to achieve ship STCP mining. The method proposed in this paper achieves accurate detection of the STCP in the AIS data by gradually eliminating redundant and worthless data and reducing the scale of mining data. Validation experiments and parameter sensitivity analysis based on real AIS datasets are conducted in the waters near the Taiwan Strait.

The method proposed in this paper is still inadequate in grid boundary processing. In future research, this is going to be improved with the voyage feature to be counted to mine stronger STCPs. This can provide some support and help for port tools used for berth allocation [41] and ship emission inventory [12] for ports. It is also expected to be applied to the detection of ship formation patterns in larger scenarios.

**Author Contributions:** Conceptualization, C.Z.and L.H.; methodology, C.Z. and G.L.; validation, C.Z. and G.L.; formal analysis, Y.W.; data curation, L.H. and G.L.; writing—original draft preparation, G.L.; writing—review and editing, C.Z.; visualization, G.L.; supervision, Y.W. and L.H.; project administration, C.Z. and Y.W.; funding acquisition, C.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is supported by the Hainan Provincial Joint Project of Sanya Yazhou Bay Science and Technology City (Grant No.2021JJLH0012), the National Science Foundation of China (Grant No.52171349), the Laboratory of Transport Pollution Control and Monitoring Technology (Grant No.2022JH-F038), the Zhejiang Provincial Science and Technology Program (Grant No. 2021C01010). And the APC was funded by the National Science Foundation of China (Grant No.52171349).

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author, Huang Liang, upon reasonable request.

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

## References

1. Dogancay, K.; Tu, Z.; Ibal, G. Research into vessel behaviour pattern recognition in the maritime domain: Past, present, and future. *Digit. Signal Process.* **2021**, *119*, 103191. [CrossRef]
2. Zheng, Y. Trajectory Data Mining: An Overview. *ACM Trans. Intell. Syst. Technol.* **2015**, *6*, 1–41. [CrossRef]
3. Wai, K.P.; Nwe Aung, T. Distance-based Clustering of Moving Objects' Trajectories from Spatiotemporal Big Data. In Proceedings of the 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, Singapore, 6–8 June 2018; pp. 567–572.
4. Murray, B.; Perera, L.P. A dual linear autoencoder approach for vessel trajectory prediction using historical AIS data. *Ocean Eng.* **2020**, *209*, 107478. [CrossRef]
5. Tang, H.; Yin, Y.; Shen, H. A model for vessel trajectory prediction based on long short-term memory neural network. *J. Mar. Eng. Technol.* **2022**, *21*, 136–145. [CrossRef]
6. Liu, T.; Ma, J. Ship Navigation Behavior Prediction Based on AIS Data. *IEEE Access* **2022**, *10*, 47997–48008. [CrossRef]
7. 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]
8. Gao, M.; Shi, G.-Y. Ship-handling behavior pattern recognition using AIS sub-trajectory clustering analysis based on the T-SNE and spectral clustering algorithms. *Ocean Eng.* **2020**, *205*, 106919. [CrossRef]
9. Duan, H.; Ma, F.; Miao, L.; Zhang, C. A semi-supervised deep learning approach for vessel trajectory classification based on AIS data. *Ocean Coast. Manag.* **2022**, *218*, 106015. [CrossRef]
10. Zaccone, R.; Martelli, M. A collision avoidance algorithm for ship guidance applications. *J. Mar. Eng. Technol.* **2020**, *19*, 62–75. [CrossRef]
11. Rong, H.; Teixeira, A.P.; Guedes Soares, C. Data mining approach to shipping route characterization and anomaly detection based on AIS data. *Ocean Eng.* **2020**, *198*, 106936. [CrossRef]
12. Toscano, D.; Murena, F.; Quaranta, F.; Mocerino, L. Assessment of the impact of ship emissions on air quality based on a complete annual emission inventory using AIS data for the port of Naples. *Ocean Eng.* **2021**, *232*, 109166. [CrossRef]
13. Bai, X.; Cheng, L.; Iris, Ç. Data-driven financial and operational risk management: Empirical evidence from the global tramp shipping industry. *Transp. Res. Part E* **2022**, *158*, 102617. [CrossRef]

14. Xing, S.; Liu, X.; He, Q.; Hampapur, A. Mining Trajectories for Spatiotemporal Analytics. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining Workshops, Brussels, Belgium, 10 December 2012; pp. 910–913.
15. Jin, J.; Zhou, W.; Jiang, B. An Overview: Maritime Spatial-Temporal Trajectory Mining. *J. Phys. Conf. Ser.* **2021**, *1757*, 012125. [CrossRef]
16. Tiantian, X.; Hong, L.; Yuqiang, L. Spatiotemporal Trajectory Gatherings Pattern Mining Method Based on R Tree Index. In Proceedings of the 2018 3rd International Conference for Convergence in Technology (I2CT), Pune, India, 6–8 April 2018; pp. 1–5.
17. Zheng, K.; Zheng, Y.; Yuan, N.J.; Shang, S. On discovery of gathering patterns from trajectories. In Proceedings of the 013 IEEE 29th International Conference on Data Engineering (ICDE), Brisbane, QL, Australia, 8–12 April 2013; pp. 242–253.
18. Li, Y.; Bailey, J.; Kulik, L. Efficient mining of platoon patterns in trajectory databases. *Data Knowl. Eng.* **2015**, *100*, 167–187. [CrossRef]
19. Celik, M.; Shekhar, S.; Rogers, J.P.; Shine, J.A. Mixed-Drove Spatiotemporal Co-Occurrence Pattern Mining. *IEEE Trans. Knowl. Data Eng.* **2008**, *20*, 1322–1335. [CrossRef]
20. Zhao, B.; Liu, X.; Jia, J.; Ji, G.; Tan, S.; Yu, Z. A Framework for Group Converging Pattern Mining using Spatiotemporal Trajectories. *Geoinformatica* **2020**, *24*, 745–776. [CrossRef]
21. Tang, L.A.; Zheng, Y.; Yuan, J.; Han, J.; Leung, A.; Hung, C.C.; Peng, W.C. On Discovery of Traveling Companions from Streaming Trajectories. In Proceedings of the 2012 IEEE 28th International Conference on Data Engineering, Arlington, VA, USA, 1–5 April 2012; pp. 186–197.
22. Puntheeranurak, S.; Shein, T.T.; Imamura, M. Efficient Discovery of Traveling Companion from Evolving Trajectory Data Stream. In Proceedings of the 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Tokyo, Japan, 23–27 July 2018; pp. 448–453.
23. Shein, T.T.; Puntheeranurak, S.; Imamura, M. Discovery of Loose Group Companion From Trajectory Data Streams. *IEEE Access* **2020**, *8*, 85856–85868. [CrossRef]
24. Aydin, B.; Angryk, R.A. *Spatiotemporal Co-Occurrence Pattern (STCOP) Mining in Briefs in Computer Science*; Springer: Cham, Switzerland, 2018; pp. 55–69.
25. Vieira, M.R.; Bakalov, P.; Tsotras, V.J. On-line discovery of flock patterns in spatio-temporal data. In Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (GIS '09); Association for Computing Machinery: New York, NY, USA, 2009; pp. 286–295.
26. Lei, B. Apriori-based Spatial Pattern Mining Algorithm for Big Data. In Proceedings of the 2020 International Conference on Urban Engineering and Management Science (ICUEMS), Zhuhai, China, 24–26 April 2020; pp. 310–313.
27. Yongmei, Z.; Sha, G.; Kuo, X.; Mengmeng, L. Mining algorithm of spatial-temporal co-occurrence pattern based on vehicle GPS trajectory. In Proceedings of the 2016 IEEE International Conference on Signal and Image Processing (ICSIP), Beijing, China, 13–15 August 2016; pp. 142–145.
28. Yadamjav, M.-E.; Bao, Z.; Zheng, B.; Choudhury, F.M.; Samet, H. Querying Recurrent Convoys over Trajectory Data. *ACM Trans. Intell. Syst. Technol.* **2020**, *11*, 1–24. [CrossRef]
29. Al-Badwi, A.; Long, Z.; Zhang, Z.; Al-Habib, M.; Al-Sabahi, K. A Novel Integrated Approach for Companion Vehicle Discovery Based on Frequent Itemset Mining on Spark. *Arab. J. Sci. Eng.* **2019**, *44*, 9517–9527. [CrossRef]
30. Zhu, M.; Liu, C.; Wang, J.; Wang, X.; Han, Y. A Service-Friendly Approach to Discover Traveling Companions Based on ANPR Data Stream. In Proceedings of the 2016 IEEE International Conference on Services Computing (SCC), San Francisco, CA, USA, 27 June–2 July 2016; pp. 171–178.
31. Wang, X.; Liu, C.; Zhu, M. Instant Traveling Companion Discovery Based on Traffic-Monitoring Streaming Data. In Proceedings of the 2016 13th Web Information Systems and Applications Conference (WISA), Wuhan, China, 23–25 September 2016; pp. 89–94.
32. Zhu, M.; Liu, C.; Wang, J.; Wang, X.; Han, Y. Instant Discovery of Moment Companion Vehicles from Big Streaming Traffic Data. In Proceedings of the 2015 International Conference on Cloud Computing and Big Data (CCBD), Shanghai, China, 4–6 November 2015; pp. 73–80.
33. Ruihong, Y.; Fei, W.; Shuhui, C. TCoD: A Traveling Companion Discovery Method Based on Clustering and Association Analysis. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019; pp. 1–7.
34. Yao, R.; Wang, F.; Chen, S.; Zhao, S. GroupSeeker: An Applicable Framework for Travel Companion Discovery from Vast Trajectory Data. *ISPRS Int. J. Geo. Inf.* **2020**, *9*, 404. [CrossRef]
35. Yu, J.; Guo, Y.; Zhu, X.; You, Y.; Xiao, D. Discovery of Travelling Companions from Trajectories with Different Sampling Rates. In Proceedings of the 4th International Conference on Computer Science and Application Engineering (CSAE 2020); Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–8.
36. Naserian, E.; Wang, X.; Xu, X.; Dong, Y. Discovery of Loose Travelling Companion Patterns from Human Trajectories. In Proceedings of the 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Sydney, NS, Australia, 12–14 December 2016; pp. 1238–1245.
37. Naserian, E.; Wang, X.; Xu, X.; Dong, Y. A Framework of Loose Travelling Companion Discovery from Human Trajectories. *IEEE Trans. Mob. Comput.* **2018**, *17*, 2497–2511. [CrossRef]

38. Wang, J.; Zhu, C.; Zhou, Y.; Zhang, W. Vessel Spatio-temporal Knowledge Discovery with AIS Trajectories Using Co-clustering. *J. Navig.* **2017**, *70*, 1383–1400. [CrossRef]
39. Zhang, Y.L.; Zhu, X.; Yang, L.J. Ships' Spatiotemporal Co-Occurrence Patterns Mining Based on Sliding Spatiotemporal Cuboid and Improved Support. In Proceedings of the 2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, 25–26 August 2018; pp. 41–46.
40. Atluri, G.; Karpatne, A.; Kumar, V. Spatio-Temporal Data Mining: A Survey of Problems and Methods. *ACM Comput. Surv.* **2018**, *51*, 1–41. [CrossRef]
41. Venturini, G.; Iris, Ç.; Kontovas, C.A.; Larsen, A. The multi-port berth allocation problem with speed optimization and emission considerations. *Transp. Res. Part D* **2017**, *54*, 142–159. [CrossRef]