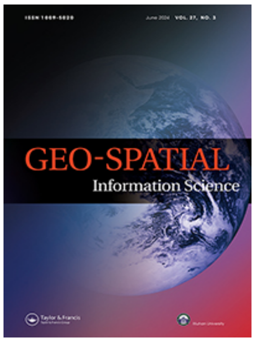


Geo-spatial Information Science

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/tgsi20

An approach for traffic pattern recognition integration of ship AIS data and port geospatial features

Gaocai Li, Xinyu Zhang, Lingling Jiang, Chengbo Wang, Ruining Huang & Zhensheng Liu

To cite this article: Gaocai Li, Xinyu Zhang, Lingling Jiang, Chengbo Wang, Ruining Huang & Zhensheng Liu (13 Feb 2024): An approach for traffic pattern recognition integration of ship AIS data and port geospatial features, Geo-spatial Information Science, DOI: 10.1080/10095020.2024.2308715

To link to this article: https://doi.org/10.1080/10095020.2024.2308715

Published online: 13 Feb 2024.

https://doi.org/10.1080/10095020.2024.2308715

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

An approach for traffic pattern recognition integration of ship AIS data and port geospatial features

Gaocai Li[a], Xinyu Zhang[a], Lingling Jiang[b], Chengbo Wang[c,a], Ruining Huang[a] and Zhensheng Liu[a]

[a]Maritime Intelligent Transportation Research Team, Navigation College, Dalian Maritime University, Dalian, China; [b]College of Environmental Science and Engineering, Dalian Maritime University, Dalian, China; [c]Department of Automation, School of Information Science and Technology, University of Science and Technology of China, Hefei, China

ABSTRACT

Recognition of ship traffic patterns can provide insights into the rules of navigation, maneuvering, and collision avoidance for ships at sea. This is essential for ensuring safe navigation at sea and improving navigational efficiency. With the popularization of the Automatic Identification System (AIS), numerous studies utilized ship trajectories to identify maritime traffic patterns. However, the current research focuses on the spatiotemporal behavioral feature clustering of ship trajectory points or segments while lacking consideration for multiple factors that influence ship behavior, such as ship static and maritime geospatial features, resulting in insufficient precision in ship traffic pattern recognition. This study proposes a ship traffic pattern recognition method that considers multi-attribute trajectory similarity (STPMTS), which considers ship static feature, dynamic feature, port geospatial feature, as well as semantic relationships between these features. First, A ship trajectory reconstruction method based on grid compression was introduced to eliminate redundant data and enhance the efficiency of trajectory similarity measurements. Subsequently, to quantify the degree of similarity of ship trajectories, a trajectory similarity measurement method is proposed that combines ship static and dynamic information with port geospatial features. Furthermore, trajectory clustering with hierarchical methods was applied based on the trajectory similarity matrix for dividing trajectories into different clusters. The quality of the similarity measurement results was evaluated by quality criterion to recognize the optimal number of ship traffic patterns. Finally, the effectiveness of the proposed method was verified using actual port ship trajectory data from the Tianjin Port of China, ranging from September to November 2016. Compared with other methods, the proposed method exhibits significant advantages in identifying traffic patterns of ships entering and leaving the port in terms of geometric features, dynamic features, and adherence to navigation rules. This study could serve as an inspiration for a comprehensive exploration of maritime transportation knowledge from multiple perspectives.

ARTICLE HISTORY
Received 24 July 2023
Accepted 17 January 2024

KEYWORDS
Ship traffic pattern; Automatic Identification System (AIS); geospatial features; semantic relationships; trajectory similarity measurement; hierarchical clustering

1. Introduction

Maritime transportation plays a crucial role in the process of trade globalization. According to the United Nations Conference on Trade and Development (UNCTAD) statistics, over 80% of global trade is conducted through maritime transportation, and within the next five years, maritime trade volume is expected to continue growing at an annual rate of 2.1% (United Nations Conference on Trade and Development 2022). With the rapid increase in international trade transport, maritime traffic safety faces severe challenges (Chen et al. 2020). Ships are the primary entities of maritime activities. Analyzing ship movement information enables the identification of ship traffic behavior patterns, which result from the collective behavior influenced by factors such as ship characteristics, crew habits, and external environmental factors. Accurate identification of ship traffic patterns contributes significantly to maritime navigation and collision avoidance (Zhang, Wang et al. 2021; Zhang et al. 2023), recognition of abnormal ship behavior (Kontopoulos, Varlamis, and Tserpes 2021; Zissis et al. 2020), maritime risk assessment (Chen et al. 2022; Zhang, Conti et al. 2021), and other aspects of maritime ship traffic safety and management (Liu, Gao et al. 2023; Pang et al. 2022), playing a crucial role for intelligent maritime applications (Liu, Yan et al. 2023).

With the widespread adoption of tracking systems such as Automatic Identification System (AIS), vast amounts of trajectory information containing ship positions, directions, and speeds are recorded and stored. It has become convenient and feasible to identify ship traffic patterns through clustering analysis of ship trajectory information. Currently, quantitative methods for identifying ship traffic patterns can be

CONTACT Xinyu Zhang zhangxy@dlmu.edu.cn

divided into two categories based on the clustering objects: trajectory points and trajectory segments. Methods that use ship trajectory points as clustering objects include k-means (Zhou, Daamen, Vellinga, and Hoogendoorn 2019), and Density-based Spatial Clustering of Applications with Noise (DBSCAN) clustering (Besse et al. 2016). Methods that use ship trajectory segments as clustering objects include spectral clustering (Gao and Shi 2020) and hierarchical clustering (Liang et al. 2021). However, these methods mainly measure the similarity of motion attribute information within ship trajectory points or segments, and classify ship trajectories accordingly. They lack the consideration of external environmental factors that influence ship movements, such as geospatial features of the water areas. Furthermore, the construction of a ship trajectory similarity measurement method that can integrate multiple attribute features is still an area requiring further research.

During maritime navigation, ship movements are influenced by factors such as the ship properties, crew habits, and geographical environment. For example, different ship types have different maneuvering performances, different operators may employ varying strategies in terms of timing, acceleration, deceleration, and turning angles to avoid collisions, and the geographic conditions of ports may lead to different route choices by vessels. Ship behaviors exhibit characteristics of randomness, diversity, and interdependency (Zhen, Shao, and Pan 2021; Zhou, Daamen, Vellinga, and Hoogendoorn 2019). Therefore, considering only the simple ship trajectory information makes it difficult to accurately and reasonably identify ship traffic patterns. It is necessary to integrate multiple attribute features to identify vessel traffic patterns. However, research on ship traffic pattern identification considering the integration of multiple attribute features is still limited, especially concerning port water areas.

To address the limitations of previous research, this paper focuses on studying ship traffic in port water areas and proposes a ship traffic pattern recognition method that considers multi-attribute trajectory similarity (STPMTS). The method consists of three modules: trajectory reconstruction, trajectory similarity measurement, and trajectory clustering. Specifically, based on the maneuvering characteristics of ships in port water areas, a grid-based compression method is proposed for trajectory reconstruction to eliminate redundant data and improve the efficiency of trajectory similarity measurement. A multi-attribute trajectory similarity measurement method is designed that integrate ship AIS data and port geographical features to address the issue of measuring similarity between trajectories with multiple attribute features related to ship movements. Hierarchical clustering is employed to classify the similarity results, and a clustering quality criterion is used to identify the optimal number of ship traffic patterns. The effectiveness of the proposed method is validated using actual ship trajectory data.

The main contributions of this research can be summarized as follows: (1) A framework for recognizing ship traffic patterns in port water areas is proposed and its effectiveness is validated using actual ship trajectory data. (2) A ship trajectory reconstruction method based on grid compression is proposed to eliminate redundant data and improve the efficiency of trajectory similarity measurement. (3) Considering the complex traffic environment in port water areas, a multi-dimensional attribute-based trajectory similarity measurement method is proposed that takes into account the relationships between attributes. (4) Based on the similarity results, hierarchical clustering is applied to cluster trajectories with multiple attribute features, and a clustering quality criterion is used to optimize the number of clusters and eliminate artificial interference.

The rest of this paper is organized as follows. Section 2 provides an overview of the research on ship traffic patterns. Section 3 presents the framework for recognizing ship traffic patterns in port water areas. Section 4 describes the methods for ship traffic pattern recognition. Validation and case studies are presented in Section 5. Finally, conclusions and future prospects are discussed. The discussion and conclusion sections are given in Section 6 and 7, respectively.

2. Related works

Research on ship traffic pattern recognition has been continuously evolving and becoming more sophisticated with variations in the format, types, and quantity of trajectory data. Currently, there are two types of trajectory-based ship traffic pattern studies: individual trajectory points and trajectory segments, respectively (Yuan et al. 2017).

For the individual trajectory points, clustering based on point distances is used to categorize ship trajectory. Yan et al. (2016) proposed a method using the DBSCAN clustering approach to identify patterns of stationary areas and frequent shipping routes of vessels, aiming to construct and analyze maritime ship traffic patterns. Zhou, Daamen, Vellinga, and Hoogendoorn (2019) introduced a ship behavior pattern recognition method in port waters based on the K-means clustering approach. They classified ships based on static features such as ship type, heading, and speed. Lee et al. (2021) applied the DBSCAN clustering algorithm to cluster trajectory points of vessels entering and exiting ports, identifying leg and turning section to assist ship maneuvering during arriving and departing from port. Rong et al. (2020) utilized the DBSCAN algorithm to cluster turning points in ship trajectories, identifying waypoints to provide vector representations of ship routes. These studies mostly focused on clustering trajectory based on geographical coordinate attributes, lacking consideration of other ship attributes such as ship type, heading, and speed. Zissis et al. (2020) proposed a method for maritime traffic pattern recognition based on a distributed framework and unsupervised learning. This method used ship types and voyage start and destination positions as the basis for voyage extraction and applied k-means clustering to ship position coordinates to generate maritime routes. Kontopoulos, Varlamis, and Tserpes (2021) presented an improved DBSCAN-based ship traffic pattern extraction method, considering ship speed, heading, and position attributes during the clustering process. However, trajectory point-based clustering methods neglect the shape characteristics of the trajectories, making it challenging to model global trajectory features and motion trends. Moreover, they do not consider external environmental factors that influence ship movements.

Clustering ship trajectory segments involves clustering continuous trajectory points to obtain more comprehensive ship traffic patterns. When clustering ship trajectory segments, it is necessary to first calculate the similarity of trajectory segments and then use clustering methods to identify ship traffic patterns. Trajectory similarity measurement methods can be broadly categorized into two types: Shape-based methods and Warping-based methods (Su et al. 2020). On the shape-based methods (Lee, Lee, and Cho 2022; Wang et al. 2021), calculate the shape differences between different trajectories using the Hausdorff distance and apply density clustering algorithms for trajectory clustering. Ma et al. (2014) developed a ship motion pattern recognition method combining the OWD distance and Spectral clustering algorithm. Shape-based similarity measurement methods typically focus only on the geometric shape of trajectories and are sensitive to trajectory noise and uneven sampling. On the Warping -based methods, Zhao and Shi (2019) calculate the similarity matrix of ship trajectory segments using dynamic time warping (DTW), combined with the DBSCAN clustering method to identify ship traffic patterns. To consider both the overall and local characteristics of trajectory segments, L. Liu et al. (2022) incorporate static and dynamic features of trajectories to calculate distance between trajectories and employ a hybrid clustering method to analyze ship traffic. Tang et al. (2021) proposed a trajectory similarity measurement method based on the geometric structure of trajectories and applied an improved FOP-OPTICS algorithm for clustering trajectory segments. Yang et al. (2022) introduced a DBTCAN clustering algorithm that utilizes the Hausdorff distance for similarity and can cluster trajectories of different lengths. In recent years, some scholars have applied deep convolutional autoencoders to learn low-dimensional features from trajectory data, providing similarity measurement methods for ship trajectory clustering (Liang et al. 2021). However, research on clustering trajectory segments still primarily relies on the trajectory data itself to calculate similarity, with little consideration for external environmental factors that affect ship movements. Overall, the current research on ship traffic pattern recognition mainly relies on ship trajectory information, such as position, heading, and speed, while lacking in-depth analysis of ship-specific attributes and the spatial-temporal binding and correlation of navigational environmental factors.

With the rapid development of Internet of Things (IoT) technologies, mobile trajectory data can now be enriched with semantic information through multiple sensors (Perera et al. 2013). New methods have been developed for measuring the similarity of multi-attribute trajectories, such as the longest common subsequence (LCSS) (Yao et al. 2017), edit distance (ED) (Fu et al. 2017). LCSS reduces the influence of noise by defining distances and matching thresholds for attributes, determining similarity values based on the relationship between distances of trajectory points and matching thresholds (Park, Jeong, and Park 2021). However, LCSS is still focused on the spatiotemporal similarity of trajectory data and assumes interdependence among all attributes. To overcome the limitations of previous approaches, Furtado et al. (2016) proposed a similarity measurement method called MSM, which can handle semantic dimensions beyond the spatiotemporal dimension of trajectories and can define specific weights based on the importance of research attributes. Nevertheless, this method overlooks the relationships between attributes. Petry et al. (2019) introduced a flexible multi-aspect trajectory similarity measurement method (MUITAS), which allows customization of distance functions and weights for different attributes and also considers semantic relationships among trajectory attributes.

To address the lack of consideration for ship-specific attributes and navigational environmental factors in current ship traffic pattern recognition research, this study proposes a ship traffic pattern recognition method that integrates multiple attribute features and utilizes a similarity measurement method for trajectory clustering that combines AIS information and port geographical features. Ship trajectories are described from three aspects: static features, dynamic features, and port geospatial features. The semantic relationships between different attribute features are taken into account, and a multi-aspect trajectory similarity measurement method is used to calculate similarity between trajectories. Hierarchical

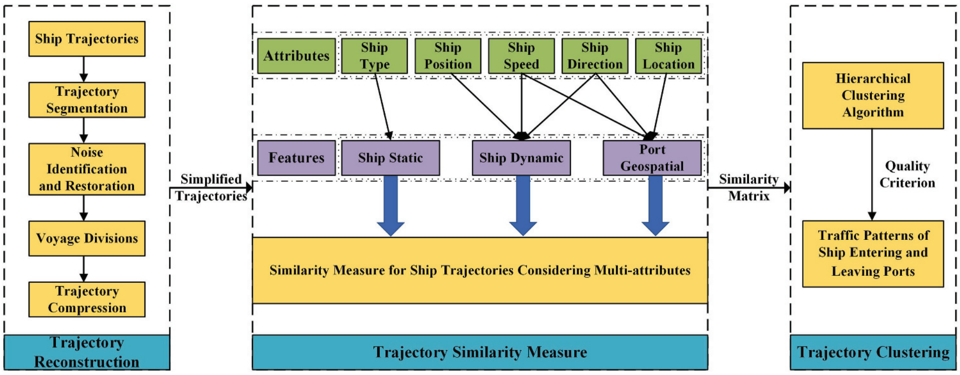


Figure 1. Framework of ship traffic pattern recognition method that considers multi-attributes trajectory similarity (STPMTS).

clustering methods are applied to identify ship traffic patterns.

3. Framework overview

To address the current issues, we propose a ship traffic pattern recognition method that considers multi-attributes trajectory similarity for ships entering and leaving port (STPMTS), as shown in Figure 1. The method consists of three modules: ship trajectory reconstruction, trajectory similarity measurement, and trajectory clustering. The trajectory reconstruction module simplifies the trajectory data by segmenting, identifying and repairing noise, and compressing trajectories to divide voyages and eliminate abnormal and redundant data, providing simplified trajectory data for trajectory similarity measurement. The ship trajectory similarity measurement module comprehensively considers the ship static and dynamic, and port geospatial feature, and the semantic relationships among them to measure the similarity matrix between ship trajectories. The ship trajectory clustering module is based on the trajectory similarity matrix and uses hierarchical clustering algorithm for aggregation and classification. The quality criteria are used to evaluate the similarity metric results and to identify the optimal number of ship traffic patterns.

4. Ship traffic pattern recognition method

4.1. Reconstruction of ship trajectories

AIS records the spatiotemporal sequence data of ship trajectories, which includes both static and dynamic data. Table 1 shows the content and update frequency of AIS information. In practical applications, due to differences in ship motion states, positioning equipment, and software design, AIS information may contain missing or erroneous data. To accurately calculate ship trajectory similarity, a ship trajectory reconstruction method was proposed based on vessel motion state, which includes three processes: trajectory segmentation, noise identification and repair, and trajectory compression. First, ship motion state estimation is used to obtain ship trajectory segments. Then, noise identification and repair of trajectory points is performed to obtain complete and accurate trajectory segments. Finally, a gridded trajectory compression method is used to remove redundant points and improve the efficiency of trajectory similarity calculation.

4.1.1. Ship trajectory segmentation

The activities of a ship typically involve three types: hoteling, maneuvering, and normal speed navigation. The hoteling refers to when the ship is engaged in activities such as cargo handling or waiting at anchor. The maneuvering refers to the ship's slow movement with the assistance of tugboats. The normal speed navigation refers to the ship's self-powered movement during navigation (Zhang et al. 2018). Figure 2 illustrates the general process of ship entering and leaving the port, which includes the three aforementioned ship activity types. Generally, the hoteling and maneuvering occur only in anchorages and basins, while the normal speed navigation can occur in any navigable water area. The update frequency of AIS dynamic information varies depending on the ship motion state. The McSMSRVD method is adopted to recognize motion states of ships (Li et al. 2022), where the inbound and outbound trajectories consist of two states: maneuvering, and normal speed navigation.

Table 1. AIS static and dynamic information (ITU 2014).

AIS information category	Content	Update frequency
Static data	MMSI, IMO, Ship type, Ship length, Ship width, Call sign, Ship name	6 min
Dynamic data	Longitude, Latitude, UTC, SOG (Speed over Ground), COG (Course over Ground), Heading, Navigation status	2 s~3 min

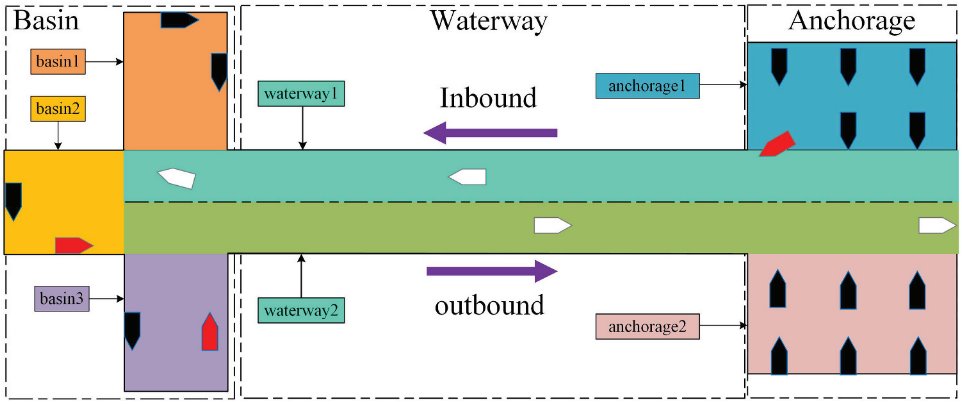


Figure 2. Schematic diagram of the flow of ships entering and leaving the port. Ships with red marks are in maneuvering state; ships with white mark are in normal speed navigation state; ships with black mark are in hoteling state.

4.1.2. Noise identification and repair

Trajectory noise identification is mainly aimed at detecting drift points in trajectory data and then fitting and restoring the trajectory segments with different motion states. For two consecutive trajectory points A and B, their distance d_{AB} can be got with Haversine formula (Chopde and Nichat 2013). Assuming the ship's motion follows uniform acceleration, the maximum length of straight path l_{AB} can be got with Newton's law of motion, that is $l_{AB} \leq v_{max} \times \Delta\ t$, in which v_{max} indicates the large value of speeds of two points and Δt is the time interval. If d_{AB} has a greater value than l_{AB}, a position drifting exists and point B should be removed. When fitting and restoring the trajectory, we consider the three types of ship activities in port waters, we adopt specific interpolation methods are applied to repair the trajectory (Zhang et al. 2018). Subsequently, we need to establish a voyage separation threshold, which is set as the minimum duration of ship's hoteling status.

4.1.3. Trajectory compression

To remove redundant data and ensure the continuity of trajectory points, we adopt a grid-based compression method to simplify the trajectory data in port waters, as shown in Equation (1). First, the study area is divided into non-repetitive $M \times N$ square grids using a geographical grid division approach, where each grid is labeled as $C_{i,j}$, with i and j representing column and row codes, respectively, as shown Figure 3(b). Then, determining the position of the point (lon_k, lat_k) in the trajectory S within the grid. Finally, we remove duplicates based on the grid location of the trajectory point, randomly selecting one point within each grid, as shown Figure 3(c). The compressed trajectory S' is obtained, where n. In this study, we set the grid cell size to 150 meters as the resolution.

$$\begin{cases} S = ((lon_1, lat_1), (lon_2, lat_2), \ldots, (lon_n, lat_n)) \\ S' = ((lon'_1, lat'_1), (lon'_2, lat'_2), \ldots, (lon'_m, lat'_m)) \end{cases} \tag{1}$$

4.2. Ship trajectory similarity measure

4.2.1. Muitas

MUITAS is a trajectory similarity measurement method that considers multiple attributes and their semantic relationships (Petry et al. 2019). One notable feature of this method is its support for both independent and dependent attributes in trajectory similarity measurement, allowing different distance functions

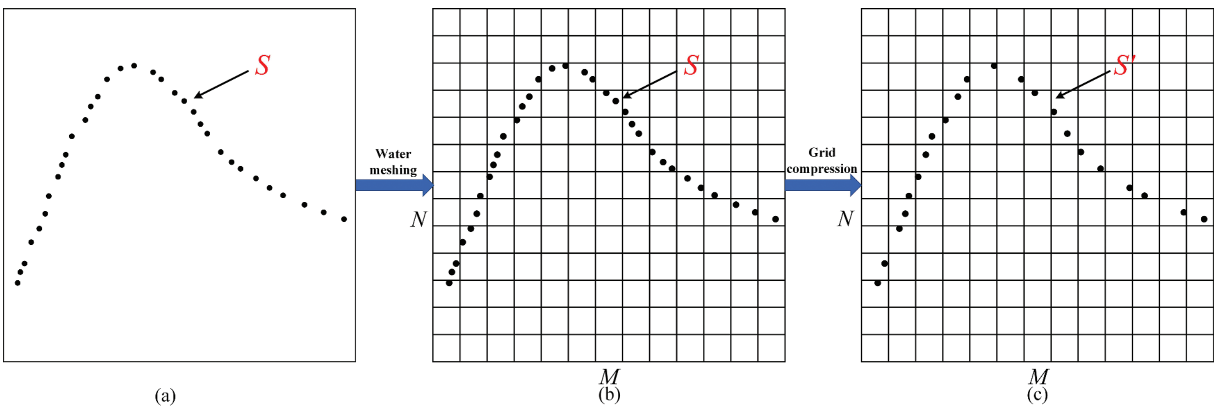


Figure 3. Schematic diagram of Grid-based trajectory compression.

and weights to be set for each attribute. Different features, distance functions, and thresholds can be set based on different application contexts.

MUITAS can be represented by a tuple, that is, $\mathbb{A} = (\mathcal{A}, \mathcal{D}, \Delta, \mathcal{F}, \mathcal{W})$, to illustrate how the similarity of multi-feature trajectories is calculated, in which $\mathcal{A} = (a_1, a_2, \ldots, a_l)$ represents a non-empty set of l attributes, $\mathcal{D} = (dist_1, dist_2, \ldots, dist_l)$ represents a non-empty set of l distance functions for the attributes, $\Delta = (\delta_1, \delta_2, \ldots, \delta_l)$ represents a non-empty set of l distance thresholds for the attributes, $\mathcal{F} = (f_1, f_2, \ldots, f_k)$ represents a non-empty set of k features, and $\mathcal{W} = (\omega_1, \omega_2, \ldots, \omega_k)$ represents a non-empty set of k feature weights. Independent attributes are defined as features with individual attributes, while attributes with semantic relationships are defined within the same feature.

Let's illustrate with a specific example. Consider two trajectories $P = <p_1, p_2, p_3>$ and $Q = <q_1, q_2, q_3>$, each with three attributes: the category and rating of visited place, representing the POI feature, and the temperature, representing the weather feature, as shown in Figure 4. In this case, if we have given attribute distance functions, distance thresholds, and feature weights, the matching score between points in $p \in P$ and $q \in Q$ can be calculated combining equations (2) and (3).

$$score(p, q) = \sum_{i=1}^{|\mathcal{F}|} (match_{f_i}(p, q) * \omega_i) \quad (2)$$

$$match_{f_i}(p, q) = \begin{cases} 1, & \text{if} \quad a_j \in f_i, dist_j(p, q) \le \delta_j \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

when calculating the matching score of multi-feature trajectory points, there may exist multiple elements in Q and $p \in P$ that have varying degrees of similarity. In this case, the sum of the highest scores of all elements

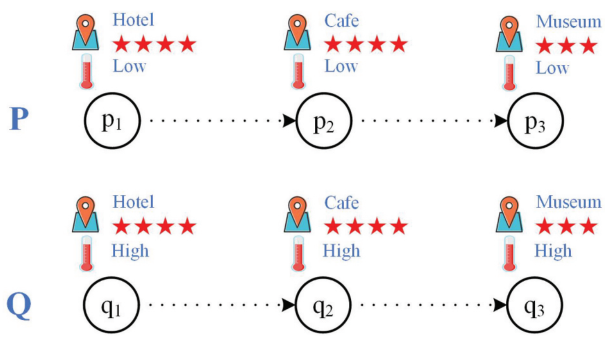


Figure 4. P and Q trajectories with multi-attributes.

in set P with any element in set Q is defined as the *parity* of set P with respect to set Q. Therefore, the *parity* can be obtained using Equation (4):

$$parity(P, Q) = \sum_{p \in P} max(\{score(p, q) | q \in Q\}) \quad (4)$$

Finally, the similarity score of P and Q is calculated by the Equation (5):

$$MUITAS = \begin{cases} 0, & \text{if} |P| = 0 \; or \; |Q| = 0 \\ \frac{parity(P,Q) + parity(Q,P)}{|P| + |Q|}, & \text{otherwise.} \end{cases} \quad (5)$$

Suppose the values of elements in the tuple are given as shown in Table 2. Table 3 presents the similarity scores between all points of both trajectories, And the *parity* $(P, Q) = parity(Q, P) = 3 \times 2/3 = 2$. The similarity score $MUITAS(P, Q) = \frac{parity(P,Q) + parity(Q,P)}{|P| + |Q|} = \frac{2+2}{3+3} = \frac{2}{3}$.

4.2.2. Multi-attribute representation of ship motion

Compared with coastal and oceanic waters, port waters have a complex traffic environment, which includes numerous anchorages, basins, and port traffic rules. In this context, ships need to follow the traffic rules established by the port to enter and exit different navigable waters to complete loading and unloading operations, as shown in Figure 2. Therefore, to fully and accurately describe ship behavior, we utilize multiple attributes such as ship type, position, speed, direction, and location to represent ship motion.

Ship type is a typical attribute of static inherent characteristics. Different types of ships have varying maneuverability. Therefore, we select ship type as a characteristic attribute to represent ship identity, and its classification is determined by the ship categories included in the AIS information. For example, if the AIS information includes container, tankers, and bulk carriers, the ship type attribute $a_1 = \{Container, Tanker, Bulk\}$.

Ship position represents the attribute of ship movement and position changes. The raw trajectory positions represented by geographic coordinates convey spatial meaning rather than ship motion. To clearly represent the characteristics of ship movement positions, based on the trajectory grid compression results in Section 4.1.3, this paper uses the row and column encoding of the grid in which the trajectory is located to represent the original trajectory coordinates, i.e. ship position attributes $a_2 = \{(Col, Row)\}$. The number of ship position attributes depends on the scale of the grid division. For example, if the study

Table 2. Features, weights, attributes, distance functions, and thresholds.

$\mathcal{F}$	$\mathcal{W}$	$\mathcal{A}$	$\mathcal{D}$	Δ
f_1	2/3	a_1 = Category	$dist_1(p, q) = 0$ if p.a_1 = q.a_1, 1 otherwise	0
		a_2 = Rating	$dist_2(p, q) = 0$ if p.a_2 = q.a_2, 1 otherwise	0
f_2	1/3	a_3 = Temperature	$dist_3(p, q) = 0$ if p.a_3 = q.a_3, 1 otherwise	0

Table 3. Scores of the points of P and Q.

$P \times Q$	q_1	q_2	q_3
p_1	2/3	0	0
p_2	0	2/3	0
p_3	0	0	2/3

area is 10 km × 10 km and the grid scale is 1 km × 1 km, a total of 10 × 10 different grids would be used to describe ship positions.

Ship speed is an attribute representing the rate of change of a ship's motion. In port waters environment, ships entering and leaving the port comply with speed limits for safe navigation. To represent the variability of ship movement speed in a more refined manner, this study adopts a segmented approach with 1 knot intervals to divide ship speed and obtain speed distribution, $v = \{(0, 1], (1, 2], \ldots, (\lfloor m \rfloor, \lceil m \rceil]\}$, where m is the maximum speed of ships in the study area. The average value of the speed segment is then taken as the speed attribute, i.e. $a_3 = \{0.5, 1.5, \ldots, (\lfloor m \rfloor + \lceil m \rceil)/2\}$. For example, if a ship's speed is 14.6 knots, its speed attribute $a_3 = 14.5$.

Ship direction represents the attribute of changes in ship movement direction. We employ the COG from the raw trajectory data to represent it. To standardize the representation of ship direction, this study utilizes a cone-based direction model to describe the ship's direction, the ship direction is divided into eight cone directions, with each sector containing a 45° angle range divided by a central axis. The angles are divided into $\theta = \{(22.5^\circ, 67.5^\circ], (67.5, 112.5], \ldots, (292.5^\circ, 337.5^\circ], (337.5^\circ, 22.5^\circ]\}$。 Consequently, semantic descriptions are obtained for the eight cone directions, which can be represented as $dir = \{Norheast, East, \ldots, Northwest, North\}$, correspondingly. For example, if a ship's direction is 85°, its direction attributes $a_4 = East$.

Ship location represents the specific functional water that a ship passes through during the navigation process. In port waters, ships usually comply with port navigation rules to move between different traffic functional water area (L. Zhang, Meng, and Fwa 2019). Therefore, ship motion exhibits distinct location characteristics. In this study, the ship's location attribute is enriched by the information of the functional area in which the ship is located (i.e. the name of the water area). The ship's location attribute is determined by the arrangement of functional water areas in port waters, including anchorages, fairways, and basins, among others. For example, if a ship is currently located in an anchorage, its location attributes $a_5 = anchorage$.

The AIS trajectory data does not directly indicate the information about the functional waters in which ships are located. Therefore, it is necessary to map the trajectory points to the corresponding water area and assign the corresponding water area names to identify the geospatial features of ships. The Even-Odd algorithm has been proven to identify whether points exist within a polygon (Hormann and Agathos 2001). Algorithm 1 provides a detailed description of the ship's location identification process functional water areas. The input consists of the ship trajectory data table T containing m trajectory points and the n distinct polygonal functional water areas WA represented by coordinate points. Firstly, a new column "Waters" is added to the trajectory data table T to indicate the location attribute of each trajectory point (line 1). Subsequently, the first trajectory point is read, and the Even-Odd algorithm is employed to identify the corresponding functional water area, modifying the "Waters" attribute value accordingly (lines 4–9). This process continues until all points in the trajectory have been examined, resulting in a saved trajectory data table with the ship's positional attributes (lines 10–12), providing the necessary data for trajectory similarity measurement.

Algorithm1: Ship location recognition in port waters.

Input: $T = \{T_i | T_i, i = 1, 2, \ldots, m\}$ (raw trajectories), $WA = \{WA_j | WA_j, j = 1, 2, \ldots, n\}$ (water areas)
Output: T_i^s (Trajectory with functional water area)
Initialize: $T_i^s \leftarrow \emptyset$

1:[1] Add a column "Waters" to the original trajectory T.
2:for $i = 1$ to m do:
3: for $j = 1$ to n do:
4: [2] Identifying the water area WA_j in which T_i.position resides using the Even-Odd algorithm and assigning the corresponding water area name.
5: if T_i.position is inside WA_j:
6: T_i.water $\leftarrow WA_j$
7: else:
8: continue
9: end for
10: [3] Assign the trajectories to T_i^s。
11: $T_i^s \leftarrow T_i$
12:end for

Table 4. The features of the three aspects of ship motion and the five attributes included.

Attribute ($\mathcal{A}$)	Feature ($\mathcal{F}$) Ship static (f_1)	Ship dynamic (f_2)	Port geospatial (f_3)
Ship Type (a_1)	✓		
Ship Position (a_2)		✓	
Ship Speed (a_3)		✓	✓
Ship Direction (a_4)		✓	✓
Ship Location (a_5)			✓

Table 5. Weights, attributes, distance function, and thresholds.

$\mathcal{W}$	$\mathcal{A}$	$\mathcal{D}$	Δ
1	a_1	$dist_1(p,q) = 0$ if p.a_1 = q.a_1, 1 otherwise	0
1	a_2	$dist_2(p,q) = 0$ if abs(p.a_2−q.a_2) = 0, 1 otherwise	1
1	a_3	$dist_3(p,q) = 0$ if abs(p.a_3−q.a) ≤ 1, 1 otherwise	1
1	a_4	$dist_4(p,q) = 0$ if p.a_4 = q.a_4, 1 otherwise	0
1	a_5	$dist_5(p,q) = 0$ if p.a_5 = q.a_5, 1 otherwise	0

4.2.3. Trajectory similarity measure

Ship motion is influenced by various factors such as inherent ship attributes and external environment, making ship motion stochastic, diverse, and coupled (Zhen, Shao, and Pan 2021; Zhou, Daamen, Vellinga, and Hoogendoorn 2019). Considering these characteristics of ship motion, this study summarizes the five attributes describing ship motion in Section 4.2.2 into three feature aspects: ship static feature, dynamic feature, and port geospatial feature, as shown in Table 4. Ship static feature include ship type attribute. Ship dynamic feature include ship position, direction, and speed attributes. Due to the constraints of port traffic environment, ships usually use similar directions and speeds when passing through the same functional water area. Based on the fact, port geospatial feature includes ship location, direction, and speed attributes. When measuring ship trajectory similarity, this study adopts the MUITAS, which considers multiple attributes and the semantic relationships between attributes, to construct the method for measuring ship trajectory similarity. For attributes described by character types, the binary distance function is employed to measure the distance between attributes, with a distance threshold set to 0. For attributes described by numerical types, the Euclidean distance function is used to measure their distance, with a threshold set to 1. To ensure fairness, each attribute is given equal weight, i.e. $w_1 : w_2 : w_3 : w_4 : w_5 = 1 : 1 : 1 : 1 : 1$. Therefore, the weights for the ship static feature, ship dynamic feature, and port geospatial feature are determined based on the number of attributes they contain, i.e. $f_1 : f_2 : f_3 = 1 : 3 : 3$. The detailed metrics for measuring ship trajectory similarity are presented in Table 5.

4.3. Clustering of ship trajectories

4.3.1. Clustering method

Once the distance matrix is obtained, we can use a clustering algorithm to divide each ship trajectory into a specific cluster. In fact, the choice of clustering method is limited by the characteristics of the trajectories. Traditional clustering methods such as *K*-means and DBSCAN are difficult to directly apply to ship trajectories with multiple descriptive features. HCA is a clustering method that only requires a distance/similarity matrix. Therefore, without loss of generality, the widely used HCA algorithm is adopted to cluster ship trajectories and extract ship traffic patterns (Fan 2019). Figure 5 displays the complete process of extracting ship traffic patterns from ship trajectories.

4.3.2. Quality criterion of cluster result

Clustering algorithms aim to group objects into homogeneous clusters that are far apart from each other. Therefore, the quality of clustering is usually evaluated by observing the variance between categories and within each category. Inspired by previous research (Besse et al. 2016), we propose using the Between-like (BC) and Within-like (WC) criteria to evaluate trajectory similarity. Ship trajectories within the same cluster should have high similarity, while trajectories from different clusters should be as far apart as possible. In other words, the WC criterion should be minimized, and the BC criterion should be maximized. Specifically, the definitions of BC and WC are given as follows:

$$\mathrm{BC} = \sum_{k=1}^{K} D(T_{\mathcal{T}}^{\mathrm{mean}}, T_{C_k}^{\mathrm{mean}}) \tag{6}$$

$$\mathrm{WC} = \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{T^i \in C_k} D(T_{C_k}^{mean}, T^i) \tag{7}$$

where $D()$ is the distance between two trajectories, C_k is the k-th cluster, K is the total number of clusters, $T_{\circ}^{mean}$ represents the average ship trajectory, and $\mathcal{T}$ represents the set of ship trajectories. However, it is mathematically challenging to compute the average ship trajectory from $\mathcal{T}$ or C_k with $k \in \{1, 2, \ldots, K\}$. Hence, we tend to select an exemplar $T_{\circ}^{ex}$ to approximate $T_{\circ}^{mean}$, for example, $T_{\mathcal{T}}^{mean} \leftarrow T_{\mathcal{T}}^{ex}$, $T_{C_k}^{mean} \leftarrow T_{\mathcal{T}}^{ex}$. Taking $\mathcal{T}$ as an example, we can define $T_{\mathcal{T}}^{ex}$ as follows:

$$T_{\mathcal{T}}^{ex} = \min_{T^i} \left\{ \sum_{j=1, j \neq i}^{M} D(T^i, T^j) \right\} \tag{8}$$

for $i \in [1, 2,]$ with M being the total number of ship trajectories in the data set $\mathcal{T}$.

Once the criteria BC and WC are obtained, we propose the following evaluation criterion (AC) to quantitatively assess the accuracy of trajectory similarity measurement, i.e. AC is defined as follows:

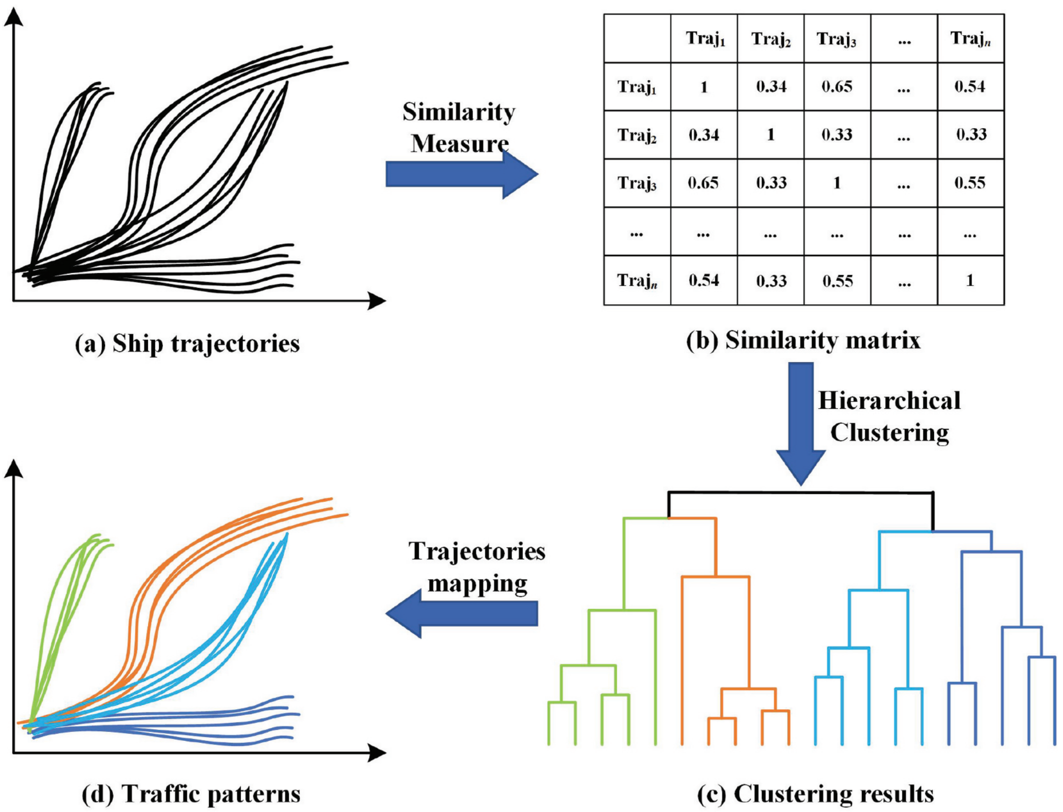


Figure 5. The complete process of extracting ship traffic patterns from ship trajectories.

$$\mathrm{AC} = \frac{BC}{BC + WC} \in (0, 1) \tag{9}$$

This criterion is essentially related to the measurement of trajectory similarity. If we can obtain accurate similarity measurements, we can generate high-quality clustering results accordingly, leading to higher AC values. Conversely, a lower AC value indicates inaccurate computation of similarity between different ship trajectories. From a theoretical perspective, as the number of clusters increases, the AC value will also increase correspondingly. Additionally, the quality criteria can also be used for selecting the number of clusters. In fact, the WC criterion value decreases with an increasing number of clusters, while the BC criterion value increases. Therefore, we should seek a balance between the criterion and the number of clusters, where adding another cluster does not significantly decrease the quality of the BC criterion.

5. Experimental and analysis

5.1. Experimental setting

To validate the effectiveness of the proposed method, a case study was conducted in Tianjin Port, located in northern China. Over 4.5 million AIS trajectory data points were collected during the period from September to November 2016. For data lacking static ship information, ship type information was supplemented through a customized web crawler and manual calibration of website resources. This study focuses on the trajectories of merchant ships. We excluded trajectories of vessels such as dredgers, barges, and fishing boats. Additionally, electronic nautical chart data was used to obtain information on the spatial geographical layout of the port's water and land areas, as shown in Figure 6. Tianjin Port has a complex navigational waterway system, including one main waterway, two boat waterways, and three warning waterways. The 11 terminal operation areas are distributed on both sides of the West and North basin. Three anchorages are located on either side of the Main Waterway.

The trajectory data was pre-processed using the trajectory data reconstruction method proposed in Section 4.1. Firstly, we set the minimum time interval between voyages to 1 hour in order to identify and segment different voyage trajectory segments (Li et al. 2022). When we observed that the trajectories of inbound or outbound voyages had very few points, set at a minimum of 20 points, indicating a significant loss of navigation data, making it challenging to accurately reconstruct the ship's trajectory. Consequently, we removed these inbound and outbound voyage trajectories, resulting in a dataset that includes trajectories of bulk carriers and container ships. Then, we identified and repaired the positional noise in the trajectory points to obtain complete and accurate trajectory segments. Considering the characteristics of the study water area, we divided it into square grid

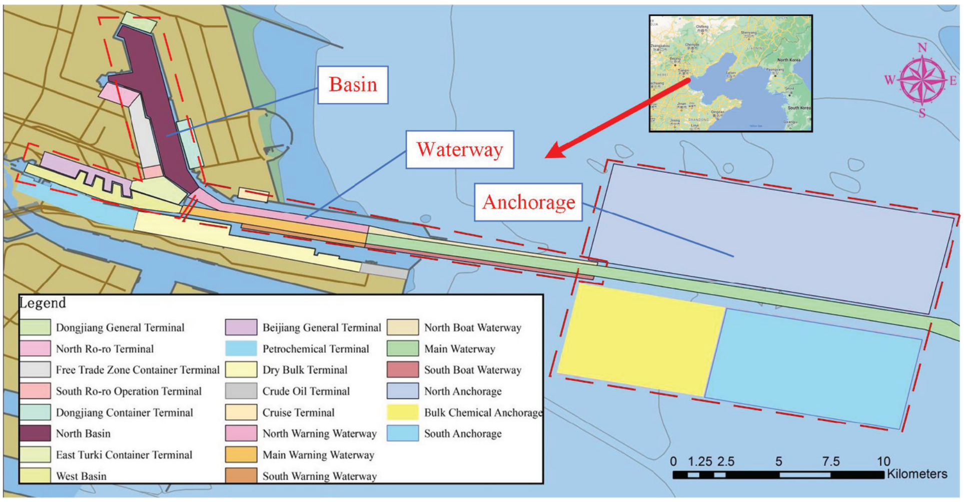


Figure 6. Spatial geographical layout of the Tianjin port's water and land areas.

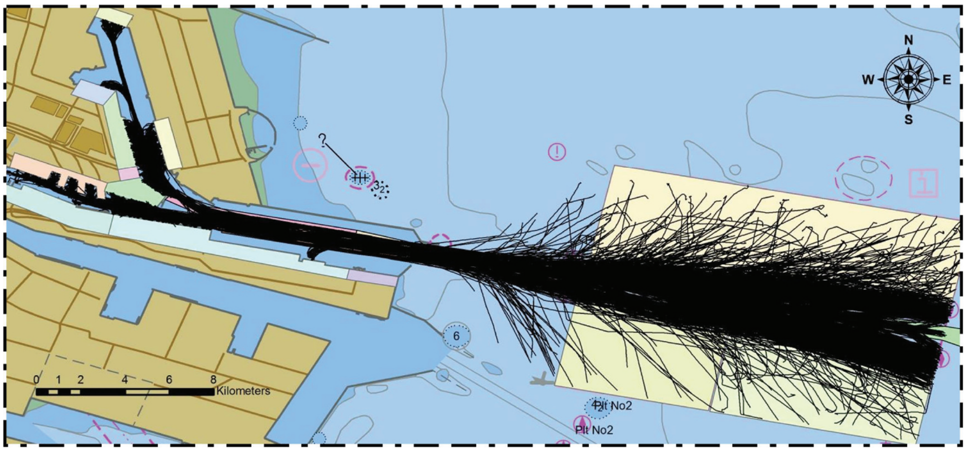


Figure 7. Visualization of the trajectory of ships entering and leaving the Tianjin Port from Sep-Nov 2016.

Table 6. Example of five attributes for three trajectory points.

NO	Ship Type	Ship Position	Ship Speed	Ship Direction	Ship Location
1	Cargo	(87246, 28871)	8.5	East	West Basin
2	Container	(87439, 28843)	14.5	East	Main Waterway
3	Cargo	(87333, 28859)	9.5	West	Main Waterway

cells with a side length of 150 meters and performed grid-based compression on the trajectory data. As a result, we obtained 1956 complete ship trajectories, with an average of 176 trajectory points per trajectory, as shown in Figure 7. Finally, we extracted five attributes, transformed and represented them, and the specific results are shown in Table 6.

5.2. Similarity calculation and quality evaluation

For the similarity measurement of multi-attribute trajectories, traditional methods such as DTW and Hausdorff, which are designed for numerical-type trajectory similarity calculation, are no longer applicable. Therefore, in order to compare and evaluate the performance of our proposed method, we will compare it with the LCSS and MSM methods, which are suitable for similarity calculation of multi-type attributes. To select a reasonable number of clusters and evaluate the clustering results, the method proposed in Section 4.3.2 is employed. The number of clusters is varied from 2 to 50 to obtain different clustering evaluation results, as shown in Figure 8.

5.3. The recognition results and analysis of ship traffic patterns

To further illustrate the effectiveness of our method, we prefer to compare the results of trajectory clustering through visualization. We designed a visualization platform that combines ArcGIS with Electronic Navigation Charts (ENCs) displaying the traffic patterns of ships entering and leaving the port. Cargo ship trajectories are shown as red lines, while container

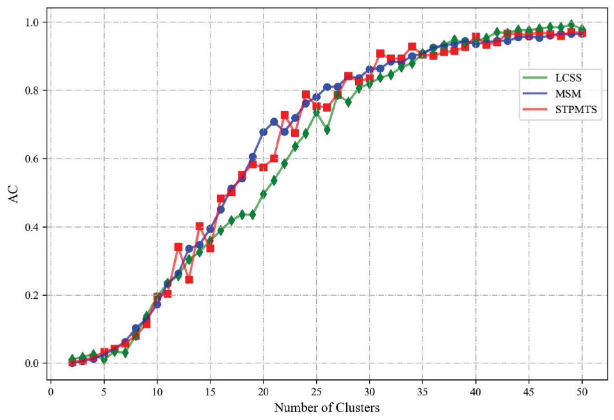
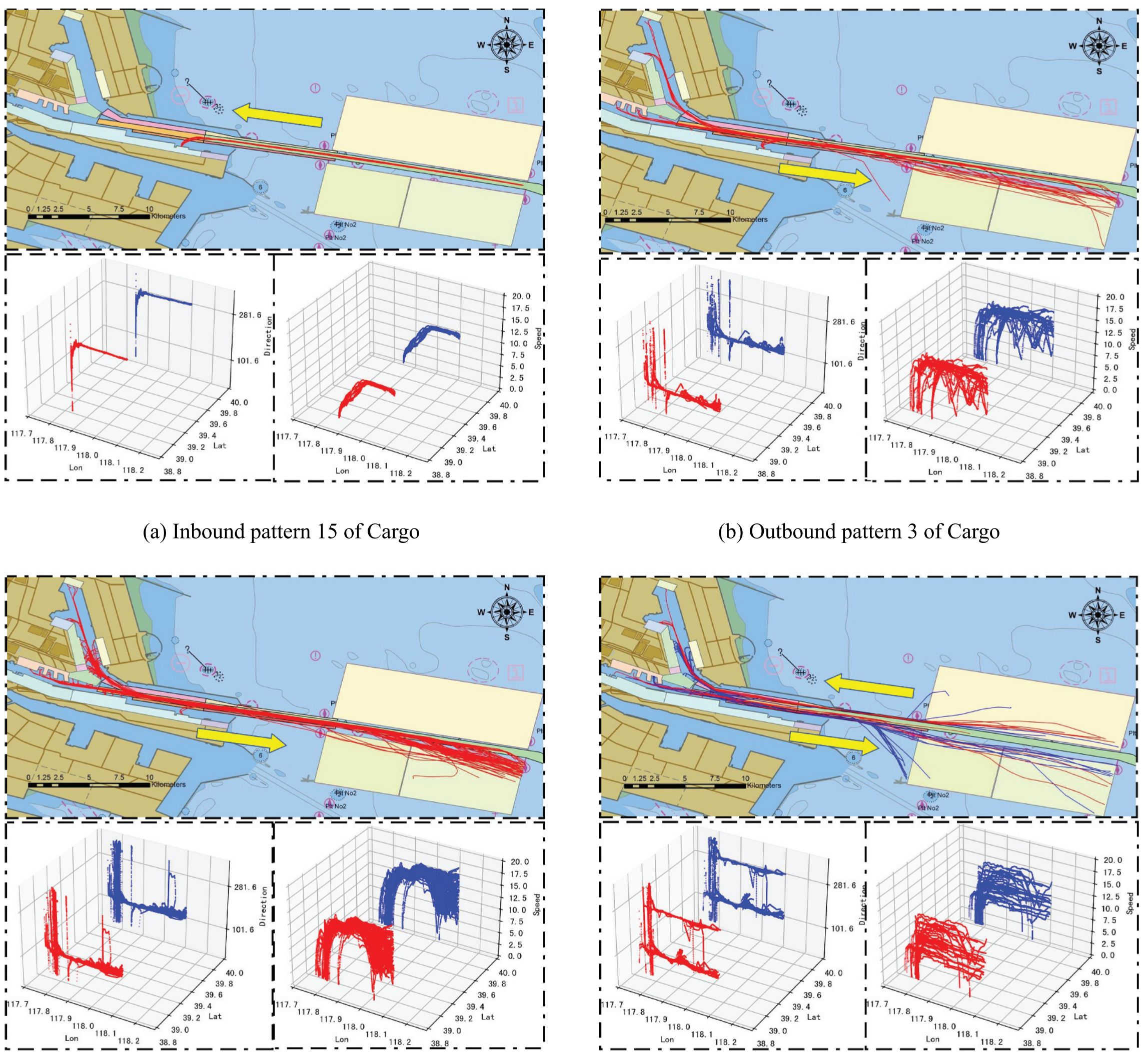


Figure 8. Results of AC criterion obtained from LCSS, MSM, STPMTS with the number of clusters.

ship trajectories are shown in blue. We use yellow arrows indicate the movement direction of trajectory clusters. Moreover, we perform spatial 3D visualization of the distribution characteristics of ship direction and speed attributes. To facilitate analysis, the 3D spatial map is projected onto the plane where y = 40. The representative inbound and outbound patterns as shown in Figures 9–11, with all the results identified by each method presented in the Appendix A. We evaluate the effectiveness of the three methods in terms of ship types, direction of inbound/outbound, geometric characteristics, behavioral characteristics, and port traffic rules, based on the visualization of the pattern results.

From the perspective of ship type and inbound/outbound direction, we summarize the results of traffic pattern recognition for the three methods, as shown in Table 7. Using the LCSS method, all trajectories are divided into 30 types of traffic patterns for cargo ships (15 inbound patterns/15 outbound patterns), 9 types for container ships (4 inbound patterns/5 outbound patterns), and 1 type for mixed traffic patterns of cargo and container ships entering and leaving the port. The MSM method divides the trajectories into 28 types of traffic patterns for cargo ships (13 inbound patterns/15 outbound patterns) and



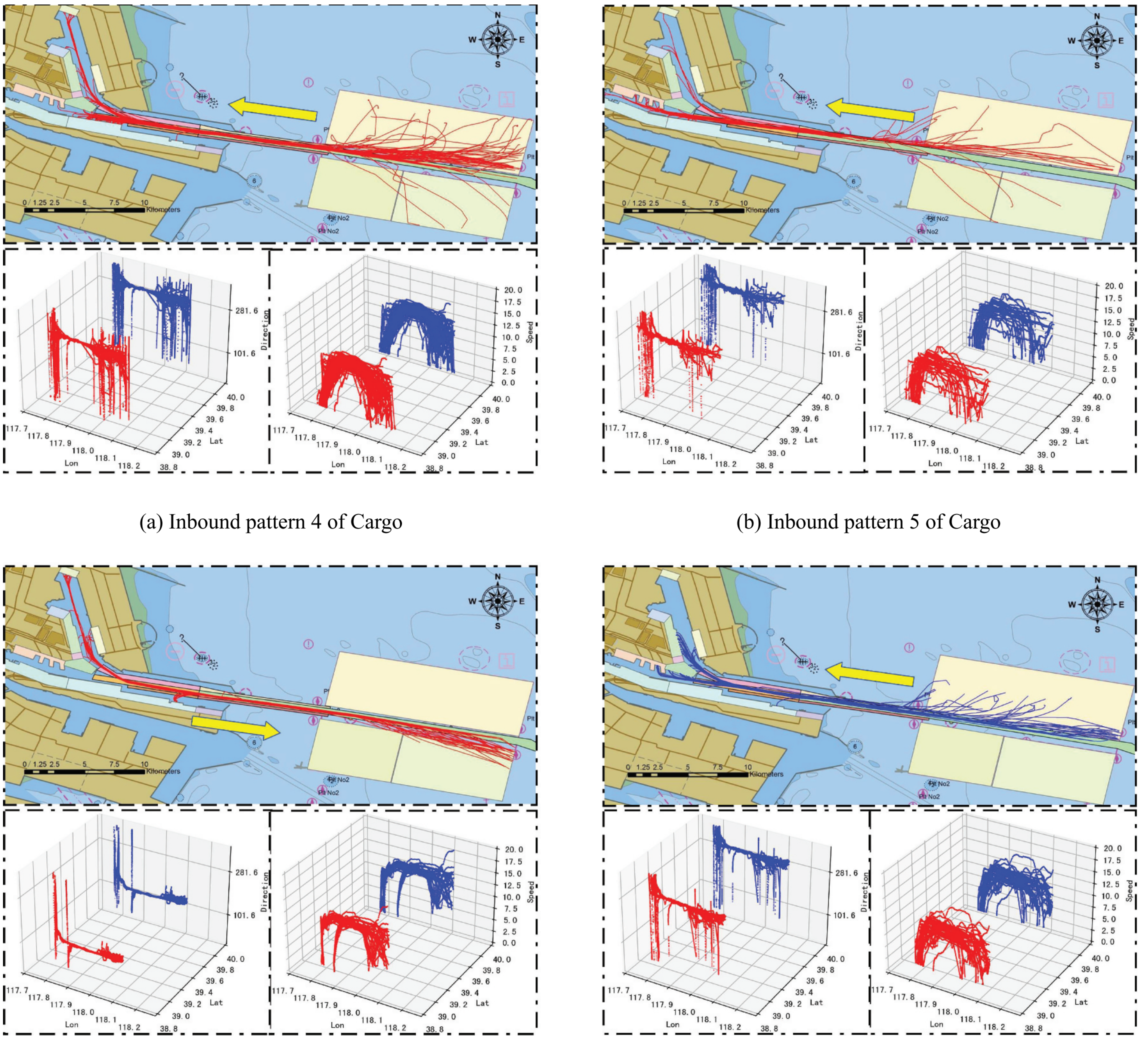
(a) Inbound pattern 15 of Cargo

(b) Outbound pattern 3 of Cargo

(c) Outbound pattern 6 of Cargo

(d) Mixed pattern

Figure 9. The recognition results of ship traffic patterns based on the LCSS method.



(a) Inbound pattern 4 of Cargo

(b) Inbound pattern 5 of Cargo

(c) Outbound pattern 1 of Cargo

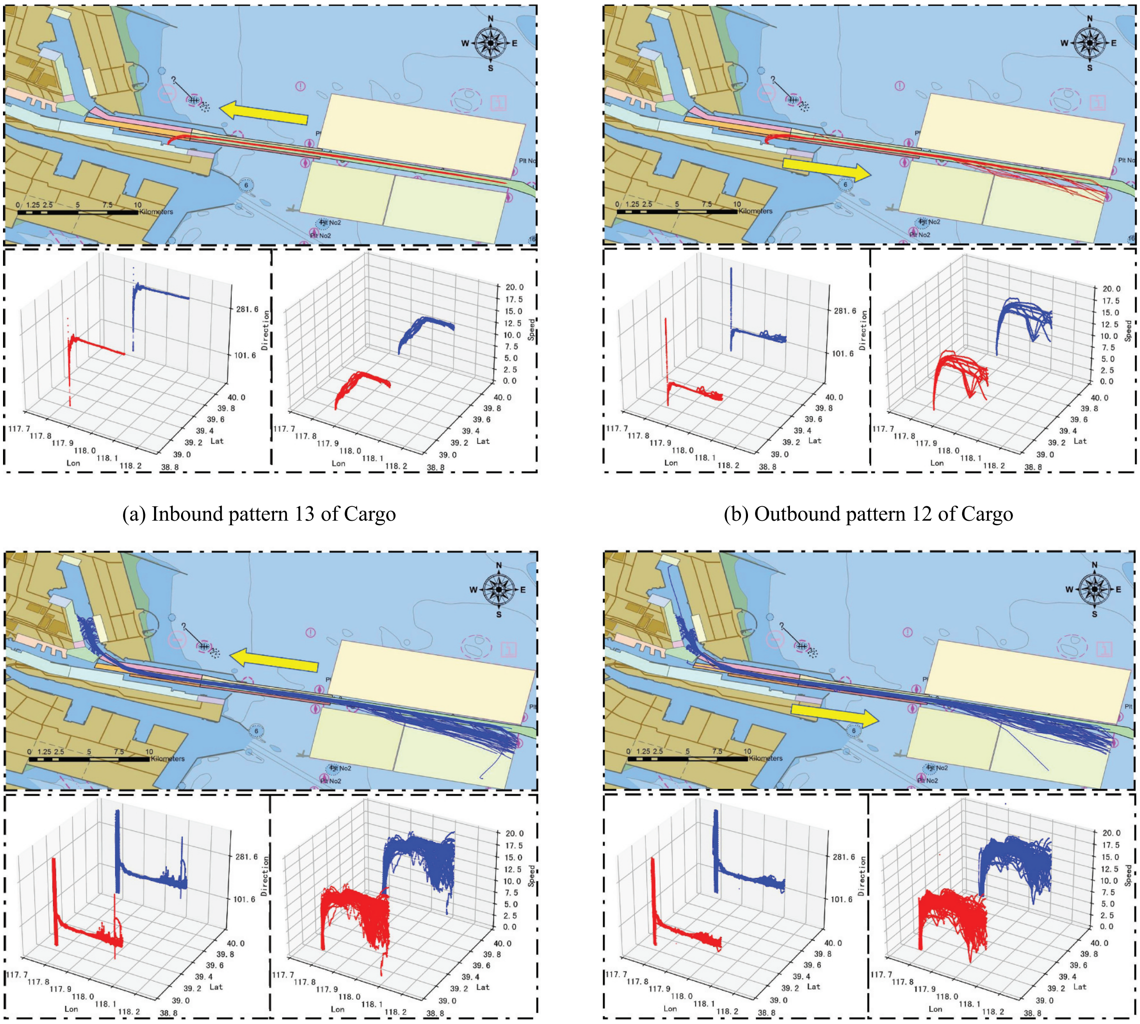
(d) Inbound pattern 1 of Container

Figure 10. The recognition results of ship traffic patterns based on the MSM method.

12 types for container ships (5 inbound patterns/7 outbound patterns). The method proposed in this paper divides the trajectories into 27 types of traffic patterns for cargo ships (13 inbound patterns/15 outbound patterns) and 13 types for container ships (6 inbound patterns/7 outbound patterns). From a comprehensive analysis of the number of patterns identified by the three methods, the LCSS method identified significantly more traffic patterns for cargo ships than the other two methods, and there were also instances of errors in identifying inbound/outbound direction and ship type. Comparing the MSM method with the STPMTS method proposed in this paper, the pattern recognition results are relatively consistent.

From the perspective of geometric characteristics, the LCSS method is prone to classifying trajectories with significant geometric differences into the same cluster, such as cargo ship inbound pattern 1. In this pattern, trajectories entering and leaving the West and North basins are observed to be recognized as the same cluster. This is because the LCSS algorithm considers two sequences to be similar if they have similar behaviors in a large part of their length, without considering the characteristics of the entire trajectory. On the other hand, the MSM and STPMTS methods can classify trajectories entering and leaving the West and North Basin well, as shown in Figure 12.

From the perspective of ship behavior characteristics, the LCSS method does not effectively differentiate ship types and direction features, as shown in Figure 9(d) depicting the mixed ship traffic pattern for inbound/outbound. In this pattern, cargo and container ships are classified into the same category, and inbound/outbound directions are not distinguished. On the other hand, the other two methods considered ship type, direction, and speed attributes and were able to better differentiate behavior feature differences. In addition, our STPMTS method takes into account the relationships between attributes and can differentiate minor differences between trajectories. For example, trajectories entering and leaving the same water area can be distinguished, as shown in Figure 13.



(a) Inbound pattern 13 of Cargo

(b) Outbound pattern 12 of Cargo

(c) Inbound pattern 4 of Container

(d) Outbound pattern 4 of Container

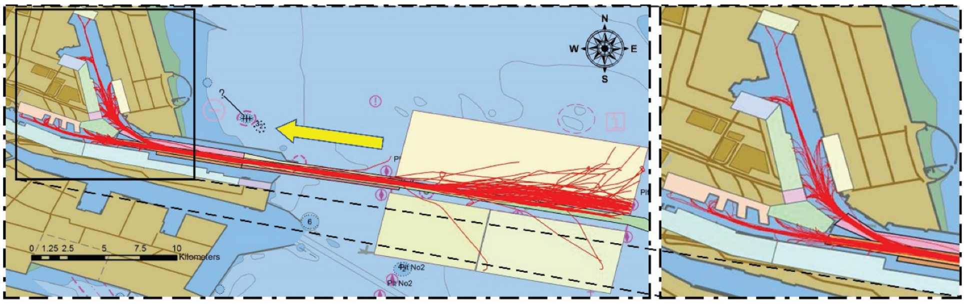
Figure 11. The recognition results of ship traffic patterns based on the STPMTS method.

Table 7. Comparing the results identified from the perspectives of ship type and direction for three methods.

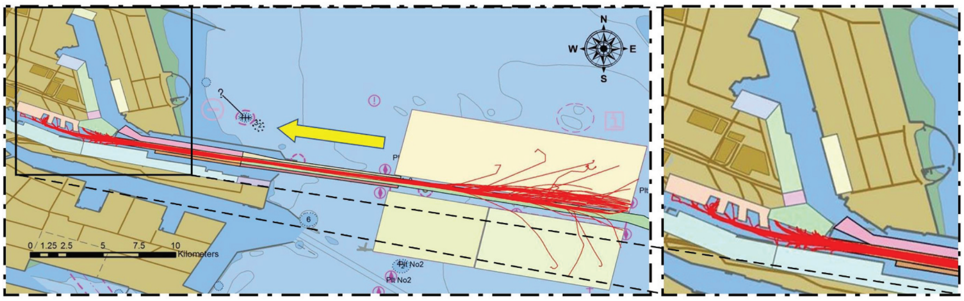
Method	Ship type	Number of patterns (Inbound/Outbound)
LCSS-based	Cargo	30(15/15)
	Container	9(4/5)
	Mixed	1
MSM-based	Cargo	28(13/15)
	Container	12(5/7)
STPMTS	Cargo	27(14/13)
	Container	13(7/6)

From the perspective of port traffic rules, the LCSS method recognizes trajectories that navigate in different waterways as the same pattern. As shown in Figure 14(a), in the outbound pattern 1 of cargo ship, the trajectories passing through the Main Waterway and the South Boat Waterway are grouped together. However, such traffic patterns clearly do not comply with the navigation rules of Tianjin Port. In contrast, the MSM and STPMTS methods proposed in this paper identify traffic patterns that are more consistent with the navigation rules of Tianjin Port. For example, in Figure 14(b), during the process of a small ship departing from the port, the vessel exits the West Basin and turns right into the South Boat Waterway and departs from the port. These vessel traffic patterns align with the navigation regulations of Tianjin Port. In Figure 14(c), during the process of a small vessel entering and berthing at the West Bain, the ship will opportunistically turn left and enter the Main Warning Waterway when passing through the North Boat Waterway.

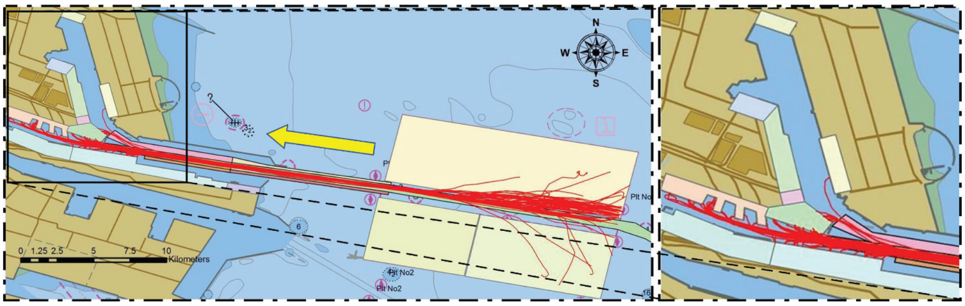
Based on the evaluation and analysis from various aspects, it can be concluded that the method proposed in this paper not only considers multiple attributes of ship traffic but also takes into account the relationships between attributes. This leads to a higher degree of similarity for similar trajectories and a greater separation of similarity for dissimilar trajectories. As a result, the classification results of ship traffic patterns are more reliable and accurate.



(a) Inbound pattern 1 of cargo.

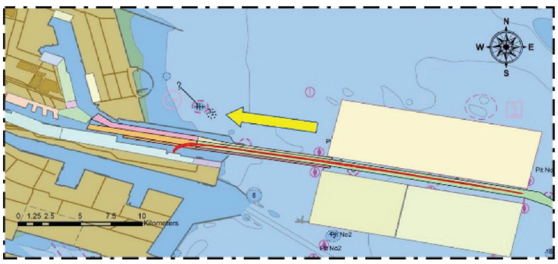


(b) Inbound pattern 11 of cargo.

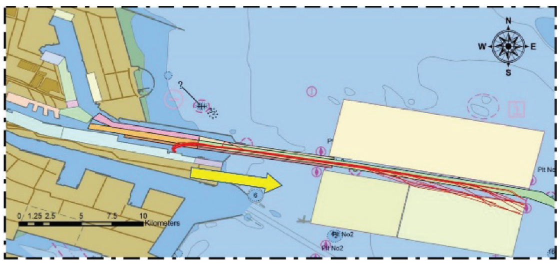


(c) Inbound pattern 7 of cargo.

Figure 12. Visual comparisons of ship trajectory clustering results from a geometric feature perspective. From top to bottom: recognition results of ship traffic patterns by (a) LCSS, (b) MSM, (c) STPMTS.



(a) Cargo ship entering Dry Bulk Terminal.

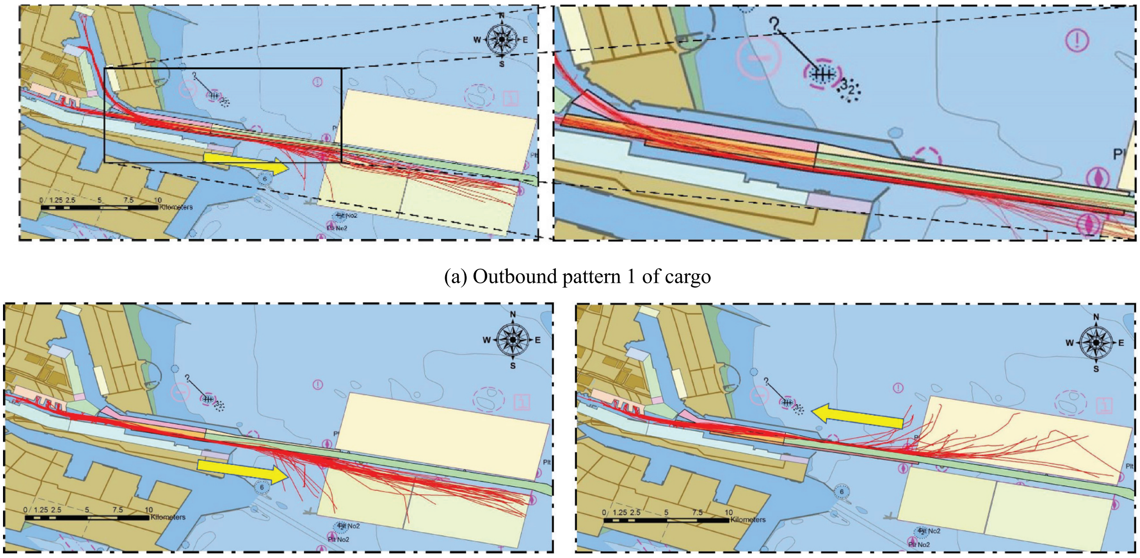


(b) Cargo ship leaving Dry Bulk Terminal.

Figure 13. Distinguish between ship trajectories entering and leaving the same terminal.

6. Discussion

This study presents an approach for ship traffic pattern recognition by integrating ship static data, dynamic data, and port geospatial features. The proposed method allows for the classification of ship trajectories related to ship entering and leaving port, as well as the analysis of their navigation characteristics. Firstly, a ship trajectory reconstruction method based on grid compression is proposed to eliminate redundant data and improve the efficiency of trajectory similarity measurement. Then, a multi-aspect trajectory similarity measurement method is employed to calculate the similarity between trajectories, taking into account the semantic relationships between multiple attribute features. Next, hierarchical clustering is used to cluster the trajectories based on the multi-attribute features, and the optimization of the number of clusters is achieved by considering



(a) Outbound pattern 1 of cargo

(b) Outbound pattern 5 of cargo

(c) Inbound pattern 6 of cargo

Figure 14. Visual comparison of three methods of identifying pattern results from perspective of port traffic rules. Recognition results of ship traffic patterns by (a) LCSS, (b) MSM, (c) STPMTS.

clustering quality criteria, aiming to find the optimal number of traffic patterns and improve the interpretability of clustering results. Finally, the proposed method is validated using real-world trajectories data, demonstrating its capability to identify and classify ship behaviors in the port water areas.

6.1. Advantages of the proposed method

Comprehensive consideration of multiple attributes and their semantic relationships: Compared to current methods, the proposed method considers various factors influencing ship behavior and establishes semantic relationships between different features. Moreover, researchers can adjust attribute-related thresholds and weights based on specific requirements to identify desired pattern results while measuring multi-attribute trajectory similarity.

Efficiency improvement through trajectory compression: During berthing and unberthing stages, ship movements are relatively slow, resulting in redundant trajectories data. Compression algorithm is proposed in this study, redundant information can be simplified. Additionally, in Section 4.2, the time complexity for measuring trajectory similarity is $O(n^2)$, where n is the length of the trajectory sequence. Adopting the proposed method, computational efficiency can be significantly improved.

User-friendliness for non-experts: Considering multiple factors influencing ship movements makes it challenging to identify ship traffic patterns accurately and reasonably. This study designs a clustering quality criterion to optimize the number of clusters, enhancing the interpretability of clustering results. Compared to traditional clustering algorithms, which require manual parameter settings based on experience, the proposed method is more user-friendly for individuals without industry expertise.

Simplicity and feasibility of the method: In terms of complexity, we compared the proposed method with another ship traffic pattern recognition method. Liu et al. (2022) proposed a classification of trajectory starting and destination locations using k-means, followed by applying DBSCAN for trajectory clustering to identify ship traffic patterns. The proposed method in this study directly utilizes hierarchical clustering to recognize ship traffic patterns by comprehensively considering various features during trajectory similarity measurement, without the need for multi-stage clustering for different features.

6.2. Limitations of the proposed method

Impact of data quality on trajectory similarity measurement accuracy. As mentioned in Section 4.2.1, the similarity of trajectories is closely related to the number of trajectory points. When there are a large number of missing or abnormal trajectory points, the accuracy of trajectory similarity calculation may be compromised, subsequently affecting ship traffic pattern recognition. Therefore, measures should be taken to reconstruct trajectories in cases of missing or abnormal trajectories. In this scenario, trajectory reconstruction can be achieved by leveraging both static and dynamic ship information (Zhang et al. 2018).

Consideration of too many attributes may decrease interpretability of pattern results. When considering attribute features and their semantic relationships, excessive attribute quantity may lead to a higher number of clusters, thus reducing the interpretability of pattern recognition results. In such cases, the following potential solutions can be considered:

(1) Feature selection based on expert knowledge or business requirements: Weighing the importance of each attribute through expert experience or selecting the most relevant attributes based on business requirements can reduce the complexity and computational burden of the clustering problem while retaining critical attribute information.
(2) Utilizing dimensionality reduction techniques to preserve key information: Transforming trajectory attribute information into a feature space and utilizing dimensionality reduction methods such as Principal Component Analysis (PCA) or t-distributed Stochastic Neighbour Embedding (t-SNE) can convert a high-dimensional feature space into a lower-dimensional representation. This helps preserve critical data information and reduce dimensionality, resulting in high-quality clustering results.
(3) Choosing suitable clustering algorithms: Selecting clustering algorithms suitable for high-dimensional data, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) or Hierarchical Clustering Analysis (HCA), can better handle multi-attribute feature data without being heavily influenced by increasing dimensions.

6.3. Imagination for real-world applications

The effectiveness of the proposed method has been validated using real ship AIS data. Ship trajectories were classified into different inbound and outbound traffic patterns, and ship traffic patterns were analyzed based on ship types, directions of inbound and outbound, behavioral characteristics, and port traffic rules. The results demonstrate that the proposed method can effectively recognize ship traffic patterns. It is anticipated that the proposed method can provide support for ship route planning, ship scheduling model construction, and decision-making in the context of ship arrivals and departures. Additionally, for different port water area environments, specific attributes can be selected based on port geography, hydrology, meteorology, and other factors to analyze maritime traffic patterns and navigation characteristics, tailored to each specific environment.

7. Conclusions and future work

This paper presents a method for recognizing port water ship traffic patterns that takes into account multiple attribute features and aims to analyze ship inbound and outbound patterns. The method involves trajectory data reconstruction to simplify trajectories and improve efficiency in trajectory similarity calculations. Integrating static ship features, dynamic features, and port geospatial features, and considering the semantic relationships between attribute features, ship trajectory similarity is computed, taking into account multiple factors that influence ship movements. The optimization of the number of clusters through clustering quality criteria enhances the interpretability of the pattern results. Overall, this method comprehensively considers multiple factors in the process of trajectory data preprocessing, trajectory similarity measurement, and trajectory clustering. Through validation using the ship trajectories data in Tianjin port, the method successfully identifies 40 ship traffic behavior patterns. Compared to other methods, it demonstrates favorable results in terms of geometric features, behavioral features, and adherence to port traffic rules, providing decision support for waterway traffic management in ports. It is speculated that this method is applicable in other maritime environments by replacing the specific traffic attributes of interest.

It is anticipated that the proposed method supports ship route planning of ship arrivals and departures, construction of ship scheduling models, and determination of scheduling schemes for ship arrivals and departures, thereby facilitating waterfront and land planning by port authorities. Future work may consider incorporating additional attribute features such as ship draft, flow velocity, etc., to obtain more meaningful conclusions. Additionally, the utilization of the traffic pattern results in ship traffic flow modeling and simulation research can be explored to evaluate port navigational efficiency. Furthermore, further research is needed to enhance the clarity in understanding the recognition results of multi-attribute feature patterns.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China [grant number 52371359] and the Dalian Science and Technology Innovation Fund [grant number 2022JJ12GX015].

Notes on contributors

Gaocai Li is currently pursuing his PhD degree in Traffic Information Engineering and Control at Dalian Maritime

University. His research interests focus on the intersection of artificial intelligence and spatiotemporal trajectory data mining, including maritime traffic knowledge extraction, ship traffic pattern recognition, and vessel behavior anomaly detection.

Xinyu Zhang is a professor and doctoral supervisor at the Navigation College of Dalian Maritime University. He is also the leader of the Maritime Intelligent Traffic Research Team. Professor Zhang holds positions as a researcher in the China Shipping 50-Person Forum and a special research fellow in the China Maritime Think Tank. He has led or participated in over 30 major projects funded by the Ministry of Science and Technology and the National Natural Science Foundation of China. He has published over 100 papers in high-level domestic and international journals, obtained more than 20 national invention patents, and received over 10 national and provincial awards for outstanding achievements in scientific research.

Lingling Jiang is a professor at the College of Environmental Science and Engineering of Dalian Maritime University. She has been engaged in research on satellite remote sensing monitoring technology for nearshore water ecological environment, remote sensing monitoring of water quality and ecological environment in coastal zones, remote sensing monitoring technology for red tide disasters, and commercial application of marine satellite technology. As the project leader, she has been in charge of several research projects, including the National Natural Science Foundation of China (NSFC) General Program, NSFC Youth Program, Liaoning Provincial Natural Science Foundation, National Oceanic Administration project, and special funding for basic scientific research in central universities.

Chengbo Wang, PhD, is a postdoctoral researcher at University of Science and Technology of China, a visiting lecturer from Liverpool John Moores University of UK. In 2023, he obtained his PhD degree in Traffic Information Engineering and Control from Dalian Maritime University, China. He conducted research primarily in the areas of reinforcement learning decision theory, autonomous ship decision planning technology, and unmanned platforms at sea. He serves as a reviewer for several prestigious international journals and conferences such as IEEE TNNLS, IEEE TITS, IEEE TASE, OE, JEET, CCS, TRB and CAC.

Ruining Huang obtained his bachelor's degree from Dalian Maritime University. He is currently pursuing a master's degree at the same institution. His research interests primarily focus on the field of artificial intelligence and big data, encompassing deep learning, smart port applications, and autonomous driving.

Zhensheng Liu obtained his bachelor's degree from Dalian Maritime University. He is currently pursuing a master's degree at the same institution. His primary research focuses on enhancing situation awareness for unmanned systems at sea, as well as autonomous ship decision-making technology. He has accumulated extensive practical experience in navigation and has actively contributed to the development of ship-assisted decision-making systems and ship behavior recognition systems.

ORCID

Gaocai Li http://orcid.org/0000-0003-0210-4594
Xinyu Zhang http://orcid.org/0000-0002-4986-4649
Chengbo Wang http://orcid.org/0000-0003-0535-1769

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. Please contact the corresponding author (i.e. X. Z. zhangxy@dlmu.edu.cn) for data sharing and cooperation.

References

Besse, P. C., B. Guillouet, J. M. Loubes, and F. Royer. 2016. "Review and Perspective for Distance-Based Clustering of Vehicle Trajectories." *IEEE Transactions on Intelligent Transportation Systems* 17 (11): 3306–3317. https://doi.org/10.1109/TITS.2016.2547641.

Chen, X., S. Liu, R. W. Liu, H. Wu, B. Han, and J. Zhao. 2022. "Quantifying Arctic Oil Spilling Event Risk by Integrating an Analytic Network Process and a Fuzzy Comprehensive Evaluation Model." *Ocean & Coastal Management* 228:106326. https://doi.org/10.1016/j.ocecoaman.2022.106326.

Chen, X., S. Wu, C. Shi, Y. Huang, Y. Yang, R. Ke, and J. Zhao. 2020. "Sensing Data Supported Traffic Flow Prediction via Denoising Schemes and ANN: A Comparison." *IEEE Sensors Journal* 20 (23): 14317–14328. https://doi.org/10.1109/JSEN.2020.3007809.

Chopde, N. R., and M. Nichat. 2013. "Landmark Based Shortest Path Detection by Using A* and Haversine Formula." *International Journal of Innovative Research in Computer and Communication Engineering* 1 (2): 298–302. http://www.ijircce.com/upload/2013/april/17_V1204030_Landmark_H.pdf.

Fan, J. 2019. "OPE-HCA: An Optimal Probabilistic Estimation Approach for Hierarchical Clustering Algorithm." *Neural Computing and Applications* 31 (7): 2095–2105. https://doi.org/10.1007/s00521-015-1998-5.

Fu, P., H. Wang, K. Liu, X. Hu, and H. Zhang. 2017. "Finding Abnormal Vessel Trajectories Using Feature Learning." *Institute of Electrical and Electronics Engineers Access* 5:7898–7909. https://doi.org/10.1109/ACCESS.2017.2698208.

Furtado, A. S., D. Kopanaki, L. O. Alvares, and V. Bogorny. 2016. "Multidimensional Similarity Measuring for Semantic Trajectories." *Transactions in GIS* 20 (2): 280–298. https://doi.org/10.1111/tgis.12156.

Gao, M., and G. Y. Shi. 2020. "Ship-Handling Behavior Pattern Recognition Using AIS Sub-Trajectory Clustering Analysis Based on the T-SNE and Spectral Clustering Algorithms." *Ocean Engineering* 205:106919. https://doi.org/10.1016/j.oceaneng.2020.106919.

Hormann, K., and A. Agathos. 2001. "The Point in Polygon Problem for Arbitrary Polygons." *Computational Geometry* 20 (3): 131–144. https://doi.org/10.1016/S0925-7721(01)00012-8.

International Telecommunication Union. 2014. "Technical Characteristics for an Automatic Identification System Using Time-Division Multiple Access in the VHF Maritime Mobile Band." Accessed December 20, 2023. https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.1371-5-201402-I!!PDF-E.pdf.

Kontopoulos, I., I. Varlamis, and K. Tserpes. 2021. "A Distributed Framework for Extracting Maritime Traffic Patterns." *International Journal of Geographical Information Science* 35 (4): 767–792. https://doi.org/10.1080/13658816.2020.1792914.

Lee, H. T., J. S. Lee, H. Yang, and I. S. Cho. 2021. "An AIS Data-Driven Approach to Analyze the Pattern of Ship Trajectories in Ports Using the DBSCAN Algorithm." *Applied Sciences* 11 (2): 799. https://doi.org/10.3390/app11020799.

Lee, J. S., H. T. Lee, and I. S. Cho. 2022. "Maritime Traffic Route Detection Framework Based on Statistical Density Analysis from AIS Data Using a Clustering Algorithm." *Institute of Electrical and Electronics Engineers Access* 10:23355–23366. https://doi.org/10.1109/ACCESS.2022.3154363.

Li, G., M. Liu, X. Zhang, C. Wang, K. H. Lai, and W. Qian. 2022. "Semantic Recognition of Ship Motion Patterns Entering and Leaving Port Based on Topic Model." *Journal of Marine Science and Engineering* 10 (12). https://doi.org/10.3390/jmse10122012.

Liang, M., R. W. Liu, S. Li, Z. Xiao, X. Liu, and F. Lu. 2021. "An Unsupervised Learning Method with Convolutional Auto-Encoder for Vessel Trajectory Similarity Computation." *Ocean Engineering* 225:108803. https://doi.org/10.1016/j.oceaneng.2021.108803.

Liu, Z., H. Gao, M. Zhang, R. Yan, and J. Liu. 2023. "A Data Mining Method to Extract Traffic Network for Maritime Transport Management." *Ocean & Coastal Management* 239:106622. https://doi.org/10.1016/j.ocecoaman.2023.106622.

Liu, J., X. Yan, C. Liu, A. Fan, and F. Ma. 2023. "Developments and Applications of Green and Intelligent Inland Vessels in China." *Journal of Marine Science and Engineering* 11 (2): 318. https://doi.org/10.3390/jmse11020318.

Liu, L., Y. Zhang, Y. Hu, Y. Wang, J. Sun, and X. Dong. 2022. "A Hybrid-Clustering Model of Ship Trajectories for Maritime Traffic Patterns Analysis in Port Area." *Journal of Marine Science and Engineering* 10 (3): 342. https://doi.org/10.3390/jmse10030342.

Ma, W., Z. Wu, J. Yang, and W. Li 2014. "Vessel Motion Pattern Recognition Based on One-Way Distance and Spectral Clustering Algorithm." In *Proceedings International Conference on Algorithms and Architectures for Parallel Processing*, 461–469. Springer. https://doi.org/10.1007/978/-3-319-11194-0_38.

Pang, X., C. Zhang, Q. Ji, Y. Chen, Z. Zhen, Y. Zhu, and Z. Yan. 2022. "Analysis of Sea Ice Conditions and Navigability in the Arctic Northeast Passage During the Summer from 2002-2021." *Geo-Spatial Information Science* 1–15. https://doi.org/10.1080/10095020.2022.2124128.

Park, J., J. Jeong, and Y. Park. 2021. "Ship Trajectory Prediction Based on Bi-LSTM Using Spectral-Clustered AIS Data." *Journal of Marine Science and Engineering* 9 (9): 1037. https://doi.org/10.3390/jmse9091037.

Perera, C., A. Zaslavsky, P. Christen, and D. Georgakopoulos. 2013. "Context Aware Computing for the Internet of Things: A Survey." *IEEE Communications Surveys & Tutorials* 16 (1): 414–454. https://doi.org/10.1109/SURV.2013.042313.00197.

Petry, L. M., C. A. Ferrero, L. O. Alvares, C. Renso, and V. Bogorny. 2019. "Towards Semantic-Aware Multiple-Aspect Trajectory Similarity Measuring." *Transactions in GIS* 23 (5): 960–975. https://doi.org/10.1111/tgis.12542.

Rong, H., A. P. Teixeira, and C. G. Soares. 2020. "Data Mining Approach to Shipping Route Characterization and Anomaly Detection Based on AIS Data." *Ocean Engineering* 198:106936. https://doi.org/10.1016/j.oceaneng.2020.106936.

Su, H., S. Liu, B. Zheng, X. Zhou, and K. Zheng. 2020. "A Survey of Trajectory Distance Measures and Performance Evaluation." *The VLDB Journal* 29 (1): 3–32. https://doi.org/10.1007/s00778-019-00574-9.

Tang, C., M. Chen, J. Zhao, T. Liu, K. Liu, H. Yan, and Y. Xiao. 2021. "A Novel Ship Trajectory Clustering Method for Finding Overall and Local Features of Ship Trajectories." *Ocean Engineering* 241:110108. https://doi.org/10.1016/j.oceaneng.2021.110108.

United Nations Conference on Trade and Development. 2022. "Review of Maritime Transport 2022." Accessed June 20, 2023. https://unctad.org/system/files/official-document/rmt2022_en.pdf.

Wang, L., P. Chen, L. Chen, and J. Mou. 2021. "Ship AIS Trajectory Clustering: An HDBSCAN-Based Approach." *Journal of Marine Science and Engineering* 9 (6): 566. https://doi.org/10.3390/jmse9060566.

Yan, W., R. Wen, A. N. Zhang, and D. Yang. 2016. "Vessel Movement Analysis and Pattern Discovery Using Density-Based Clustering Approach." In *2016 IEEE International Conference on Big Data (Big Data)*, 3798–3806. Washington, USA, December 5-8. https://doi.org/10.1109/BigData.2016.7841051.

Yang, J., Y. Liu, L. Ma, and C. Ji. 2022. "Maritime Traffic Flow Clustering Analysis by Density Based Trajectory Clustering with Noise." *Ocean Engineering* 249:111001. https://doi.org/10.1016/j.oceaneng.2022.111001.

Yao, D., C. Zhang, Z. Zhu, J. Huang, and J. Bi. 2017. "Trajectory Clustering via Deep Representation Learning." In *2017 International Joint Conference on Neural Networks (IJCNN)*, 3880–3887. Anchorage, USA: IEEE, May 14-19. https://doi.org/10.1109/IJCNN.2017.7966345.

Yuan, G., P. Sun, J. Zhao, D. Li, and C. Wang. 2017. "A Review of Moving Object Trajectory Clustering Algorithms." *Artificial Intelligence Review* 47 (1): 123–144. https://doi.org/10.1007/s10462-016-9477-7.

Zhang, L., Q. Meng, and T. F. Fwa. 2019. "Big AIS Data Based Spatial-Temporal Analyses of Ship Traffic in Singapore Port Waters." *Transportation Research Part E: Logistics & Transportation Review* 129:287–304. https://doi.org/10.1016/j.tre.2017.07.011.

Zhang, L., Q. Meng, Z. Xiao, and X. Fu. 2018. "A Novel Ship Trajectory Reconstruction Approach Using AIS Data." *Ocean Engineering* 159:165–174. https://doi.org/10.1016/j.oceaneng.2018.03.085.

Zhang, M., F. Conti, H. Le Sourne, D. Vassalos, P. Kujala, D. Lindroth, and S. Hirdaris. 2021. "A Method for the Direct Assessment of Ship Collision Damage and Flooding Risk in Real Conditions." *Ocean Engineering* 237:109605. https://doi.org/10.1016/j.oceaneng.2021.109605.

Zhang, M., P. Kujala, M. Musharraf, J. Zhang, and S. Hirdaris. 2023. "A Machine Learning Method for the Prediction of Ship Motion Trajectories in Real Operational Conditions." *Ocean Engineering* 283:114905. https://doi.org/10.1016/j.oceaneng.2023.114905.

Zhang, X., C. Wang, L. Jiang, L. An, and R. Yang. 2021. "Collision-Avoidance Navigation Systems for Maritime Autonomous Surface Ships: A State of the Art Survey." *Ocean Engineering* 235:109380. https://doi.org/10.1016/j.oceaneng.2021.109380.

Zhao, L., and G. Shi. 2019. "A Trajectory Clustering Method Based on Douglas-Peucker Compression and Density for Marine Traffic Pattern Recognition." *Ocean Engineering* 172:456–467. https://doi.org/10.1016/j.oceaneng.2018.12.019.

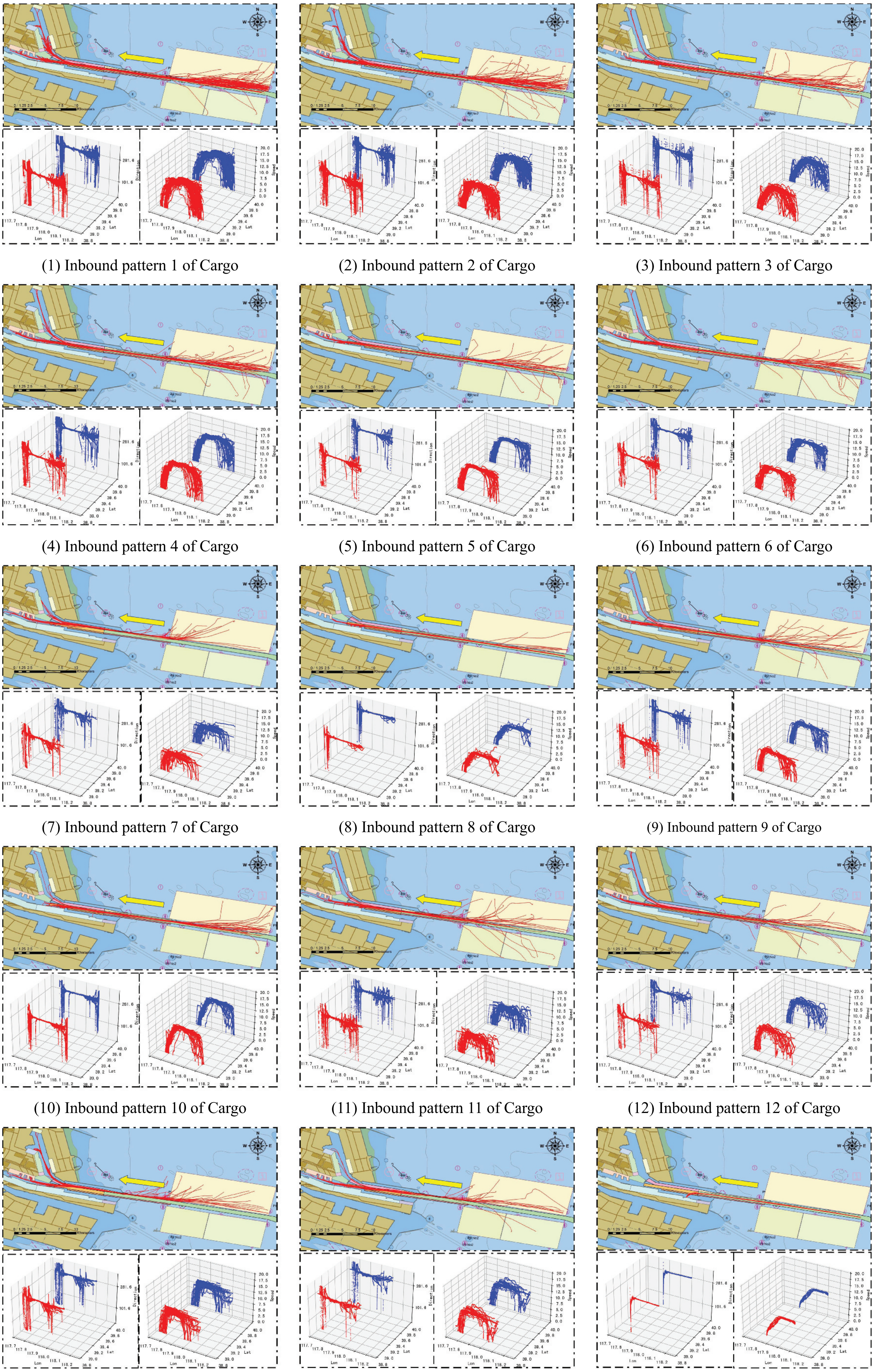
Zhen, R., Z. P. Shao, and J. C. Pan. 2021. "Advance in Character Mining and Prediction of Vessel Behavior Based on AIS Data." *Journal of Geo-Information Science* 23 (12): 2111–2127. https://doi.org/10.12082/dqxxkx.2021.210495.

Zhou, Y., W. Daamen, T. Vellinga, and S. Hoogendoorn. 2019. "Review of Maritime Traffic Models from Vessel Behavior Modeling Perspective." *Transportation Research Part C: Emerging Technologies* 105:323–345. https://doi.org/10.1016/j.trc.2019.06.004.

Zhou, Y., W. Daamen, T. Vellinga, and S. P. Hoogendoorn. 2019. "Ship Classification Based on Ship Behavior Clustering from AIS Data." *Ocean Engineering* 175:176–187. https://doi.org/10.1016/j.oceaneng.2019.02.005.

Zissis, D., K. Chatzikokolakis, G. Spiliopoulos, and M. Vodas. 2020. "A Distributed Spatial Method for Modeling Maritime Routes." *Institute of Electrical and Electronics Engineers Access* 8:47556–47568. https://doi.org/10.1109/ACCESS.2020.2979612.

Appendix



(1) Inbound pattern 1 of Cargo (2) Inbound pattern 2 of Cargo (3) Inbound pattern 3 of Cargo

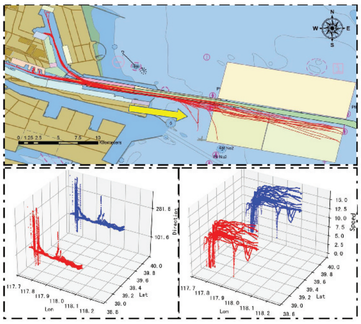
(4) Inbound pattern 4 of Cargo (5) Inbound pattern 5 of Cargo (6) Inbound pattern 6 of Cargo

(7) Inbound pattern 7 of Cargo (8) Inbound pattern 8 of Cargo (9) Inbound pattern 9 of Cargo

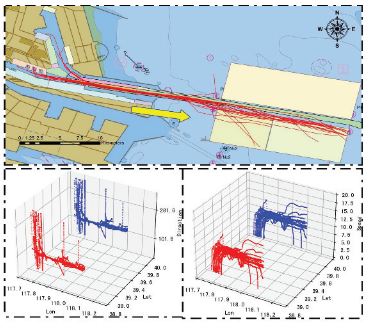
(10) Inbound pattern 10 of Cargo (11) Inbound pattern 11 of Cargo (12) Inbound pattern 12 of Cargo

(13) Inbound pattern 13 of Cargo (14) Inbound pattern 14 of Cargo (15) Inbound pattern 15 of Cargo

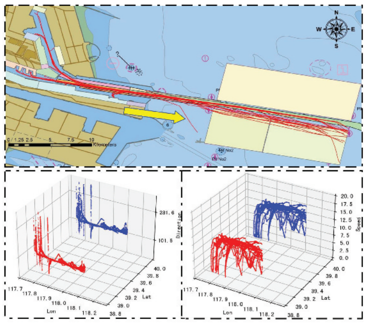
Figure A1a. The recognition results of ship traffic patterns based on the LCSS method.



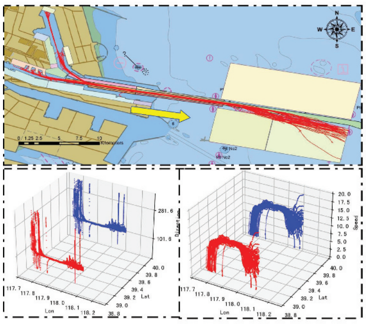
(16) Outbound pattern 1 of Cargo



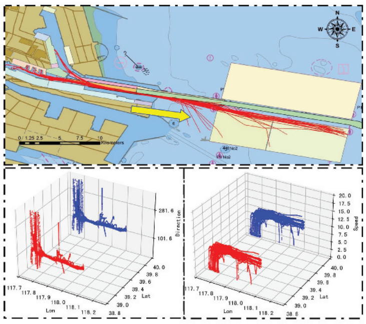
(17) Outbound pattern 2 of Cargo



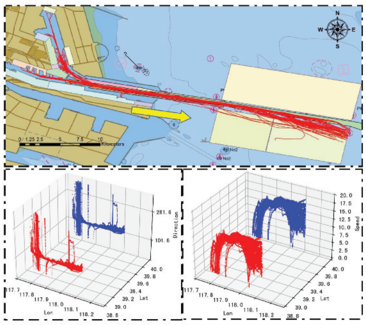
(18) Outbound pattern 3 of Cargo



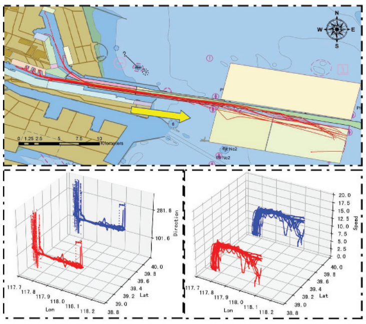
(19) Outbound pattern 4 of Cargo



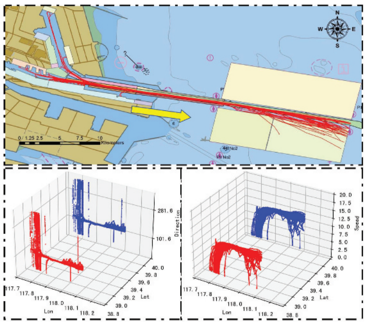
(20) Outbound pattern 5 of Cargo



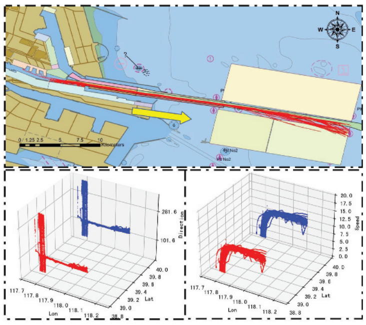
(21) Outbound pattern 6 of Cargo



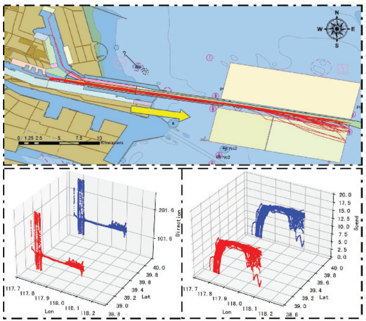
(22) Outbound pattern 7 of Cargo



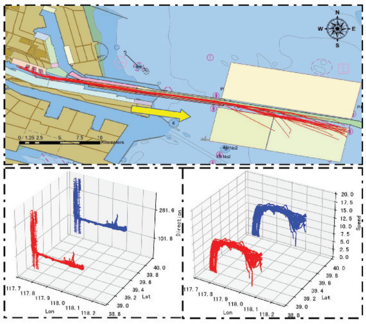
(23) Outbound pattern 8 of Cargo



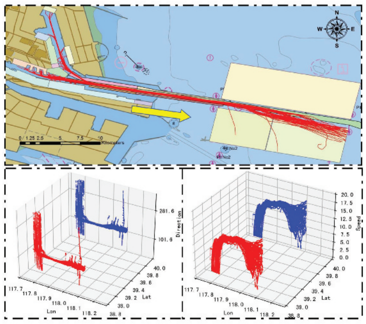
(24) Outbound pattern 9 of Cargo



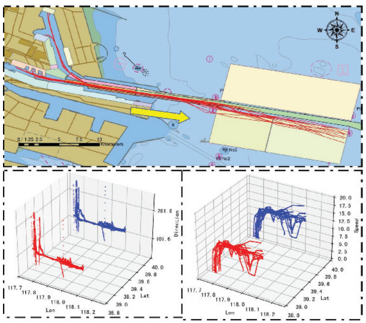
(25) Outbound pattern 10 of Cargo



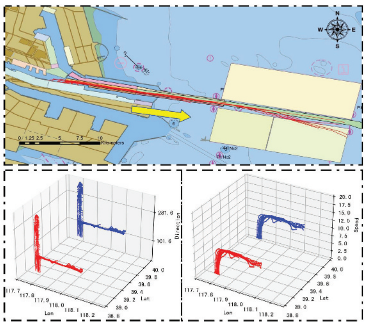
(26) Outbound pattern 11 of Cargo



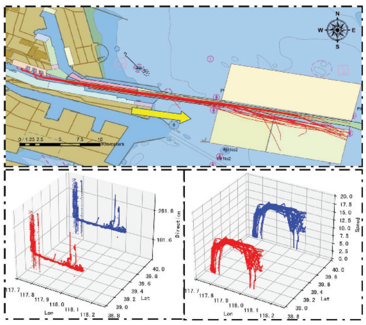
(27) Outbound pattern 12 of Cargo



(28) Outbound pattern 13 of Cargo

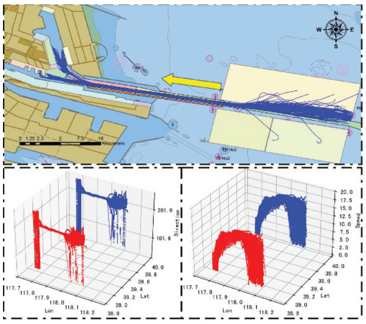


(29) Outbound pattern 14 of Cargo

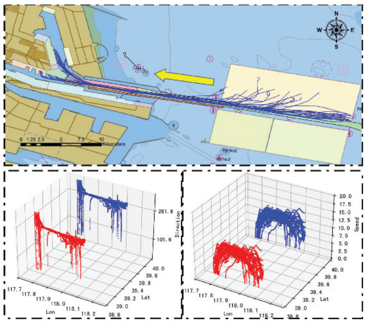


(30) Outbound pattern 15 of Cargo

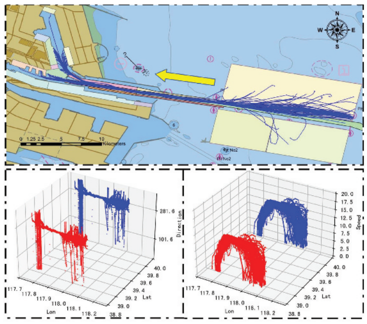
Figure A1b.



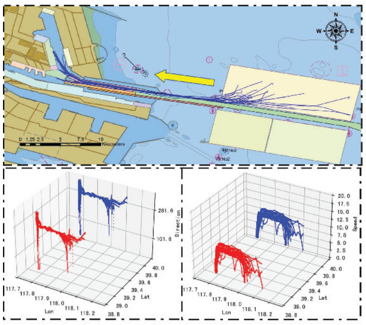
(31) Inbound pattern 1 of Container



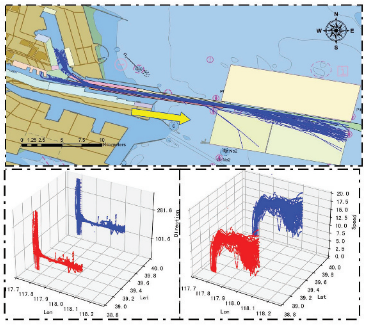
(32) Inbound pattern 2 of Container



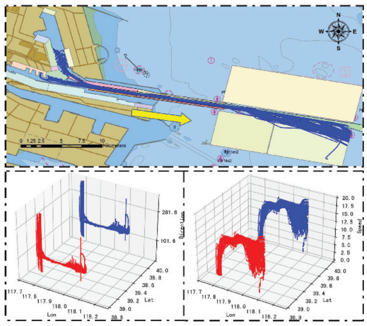
(33) Inbound pattern 3 of Container



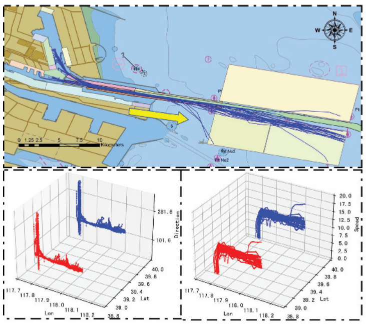
(34) Inbound pattern 4 of Container



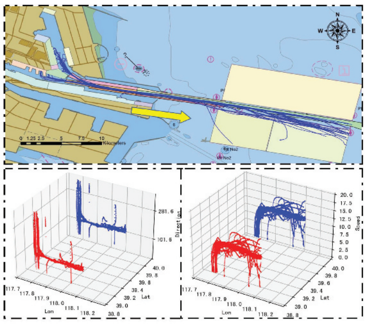
(35) Outbound pattern 1 of Container



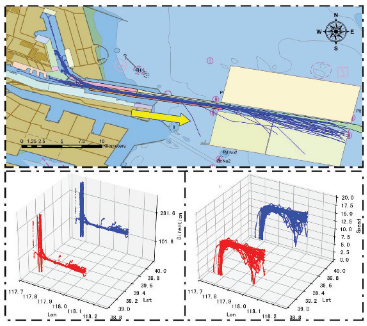
(36) Outbound pattern 2 of Container



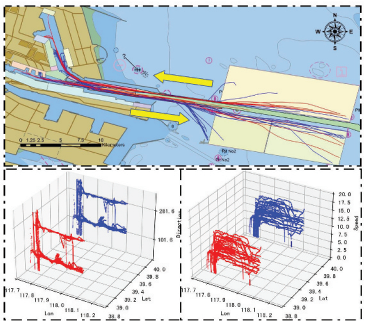
(37) Outbound pattern 3 of Container



(38) Outbound pattern 4 of Container

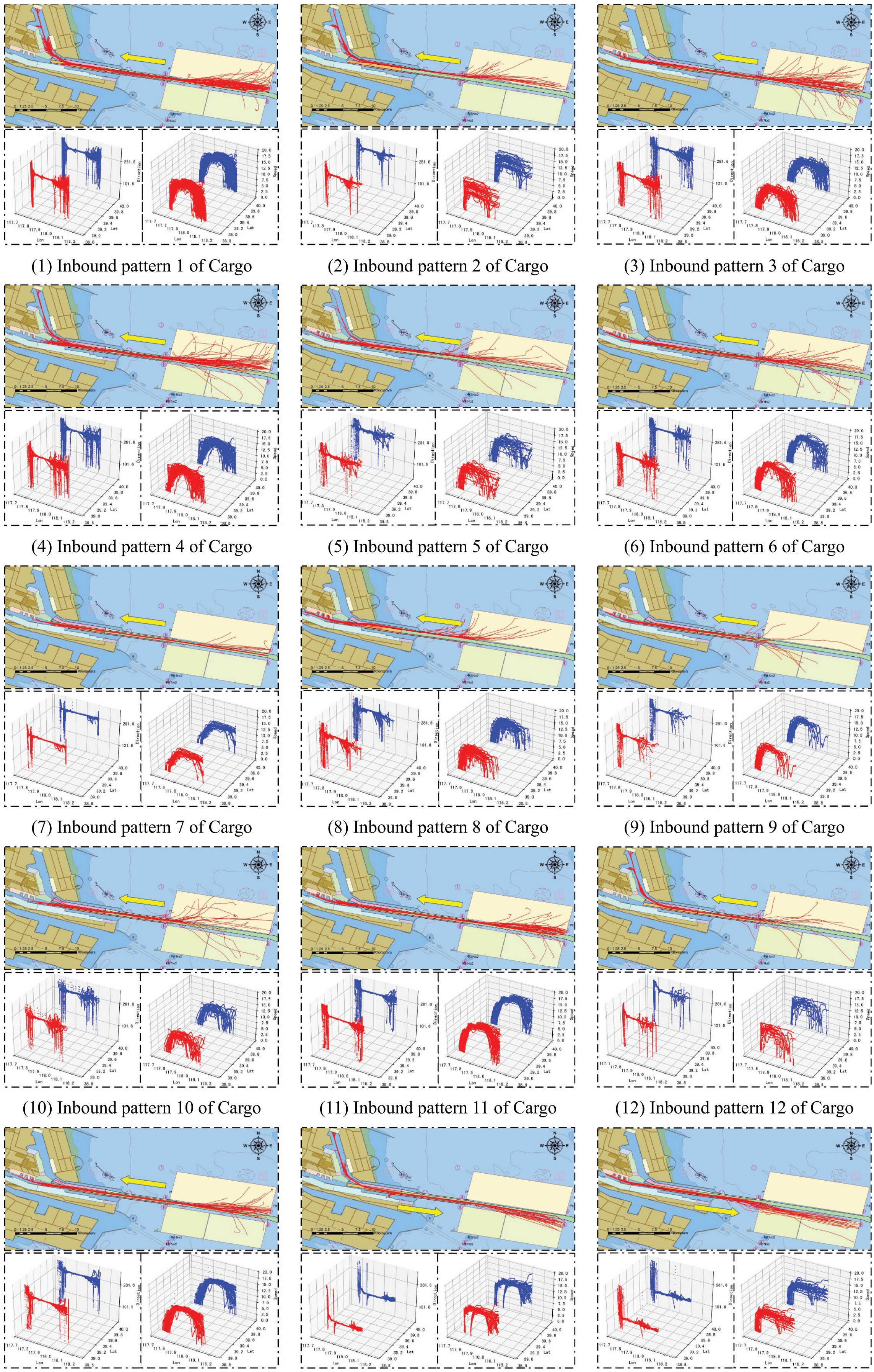


(39) Outbound pattern 5 of Container



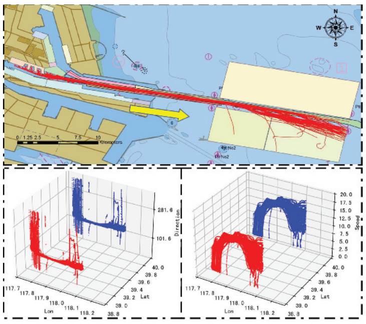
(40) Mixed pattern

Figure A1c.

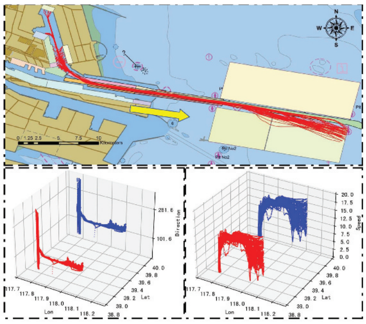


(1) Inbound pattern 1 of Cargo (2) Inbound pattern 2 of Cargo (3) Inbound pattern 3 of Cargo
(4) Inbound pattern 4 of Cargo (5) Inbound pattern 5 of Cargo (6) Inbound pattern 6 of Cargo
(7) Inbound pattern 7 of Cargo (8) Inbound pattern 8 of Cargo (9) Inbound pattern 9 of Cargo
(10) Inbound pattern 10 of Cargo (11) Inbound pattern 11 of Cargo (12) Inbound pattern 12 of Cargo
(13) Inbound pattern 13 of Cargo (14) Outbound pattern 1 of Cargo (15) Outbound pattern 2 of Cargo

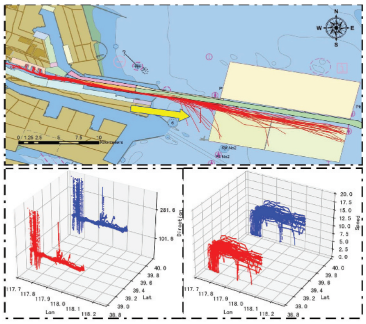
Figure A2a. The recognition results of ship traffic patterns based on the MSM method.



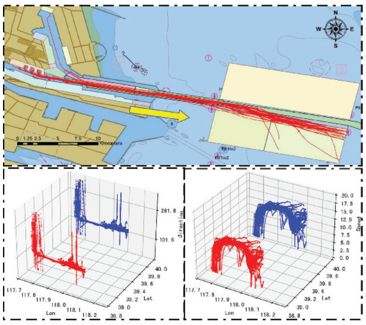
(16) Outbound pattern 3 of Cargo



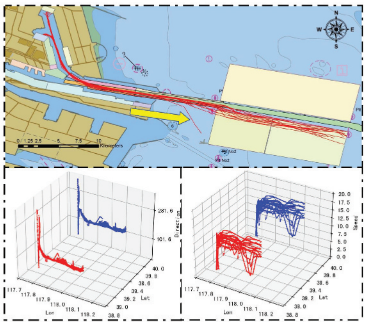
(17) Outbound pattern 4 of Cargo



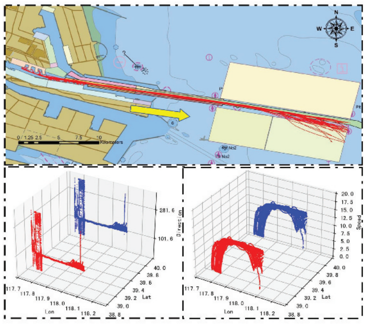
(18) Outbound pattern 5 of Cargo



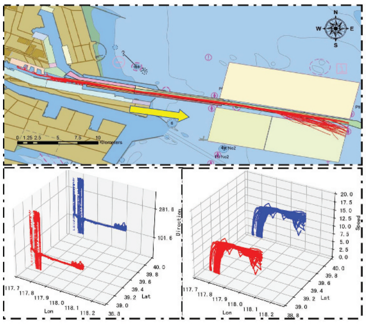
(19) Outbound pattern 6 of Cargo



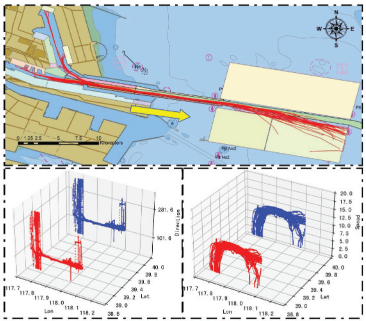
(20) Outbound pattern 7 of Cargo



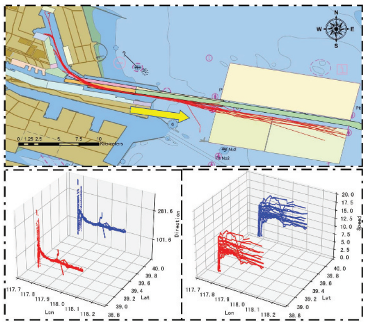
(21) Outbound pattern 8 of Cargo



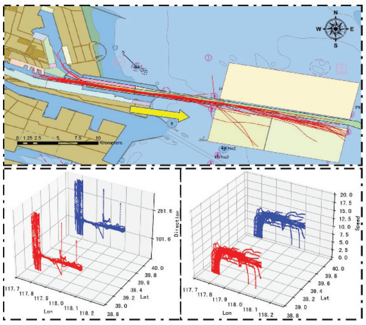
(22) Outbound pattern 9 of Cargo



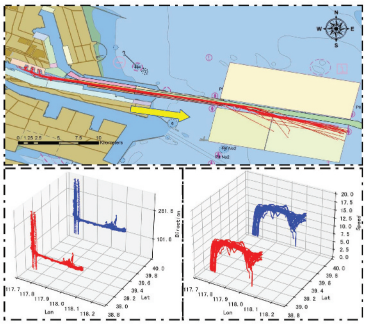
(23) Outbound pattern 10 of Cargo



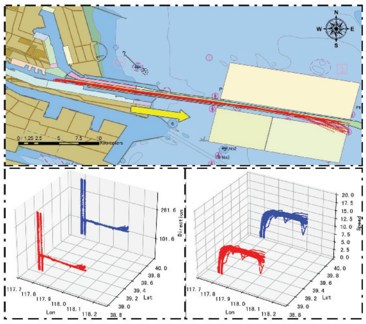
(24) Outbound pattern 11 of Cargo



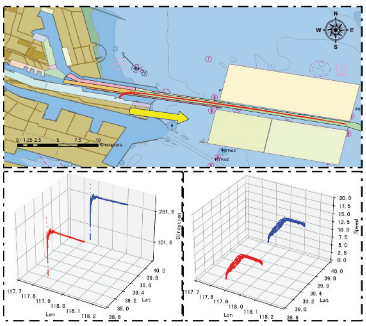
(25) Outbound pattern 12 of Cargo



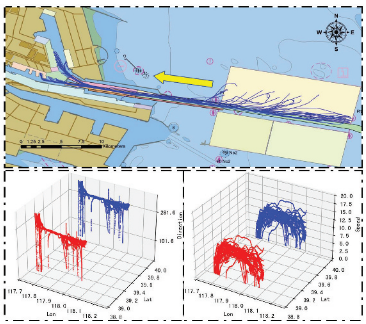
(26) Outbound pattern 13 of Cargo



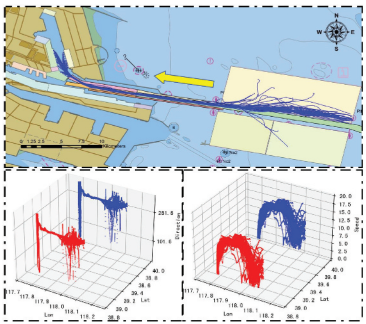
(27) Outbound pattern 14 of Cargo



(28) Outbound pattern 15 of Cargo

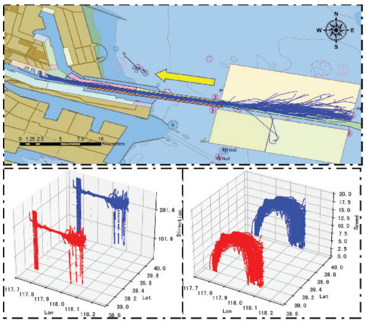


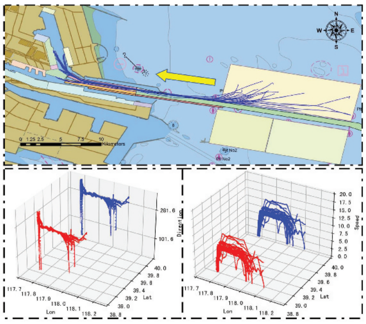
(29) Inbound pattern 1 of Container

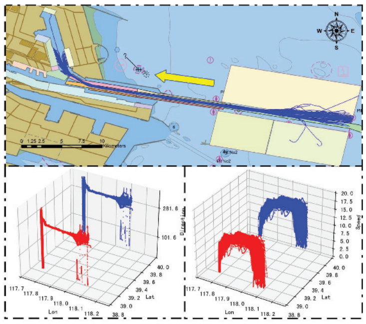


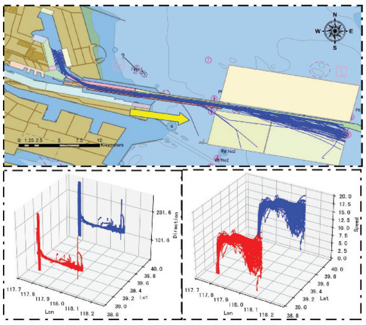
(30) Inbound pattern 2 of Container

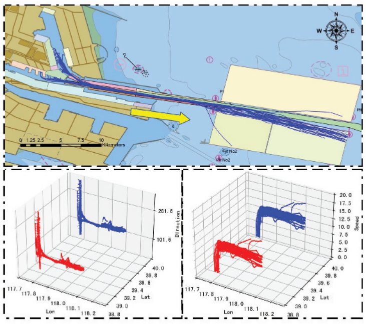
Figure A2b.

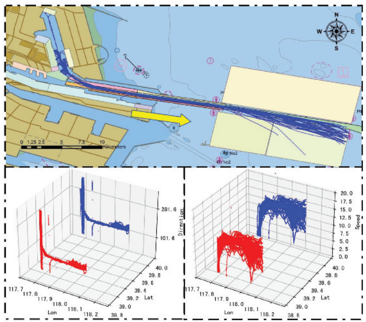

(31) Inbound pattern 3 of Container

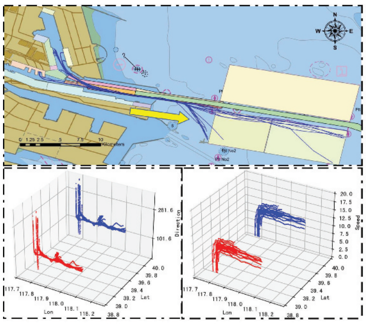

(32) Inbound pattern 4 of Container

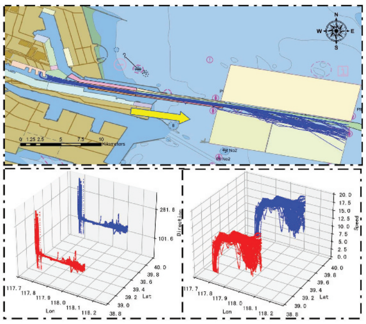

(33) Inbound pattern 5 of Container

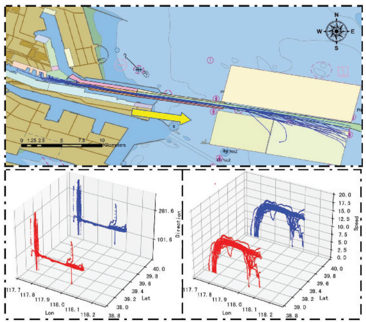

(34) Outbound pattern 1 of Container


(35) Outbound pattern 2 of Container


(36) Outbound pattern 3 of Container


(37) Outbound pattern 4 of Container


(38) Outbound pattern 5 of Container


(39) Outbound pattern 6 of Container

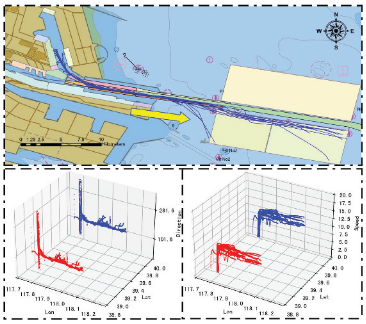
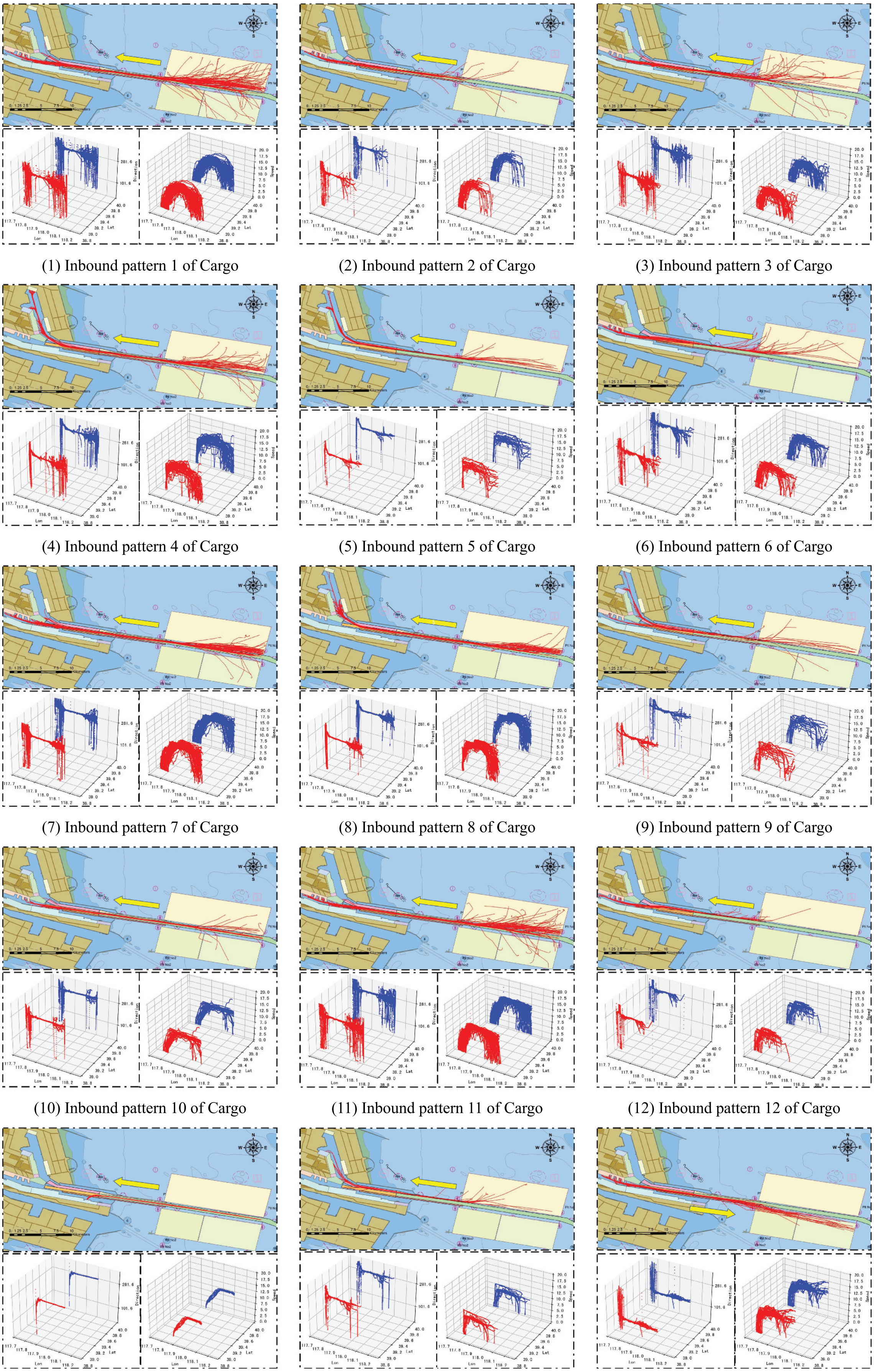
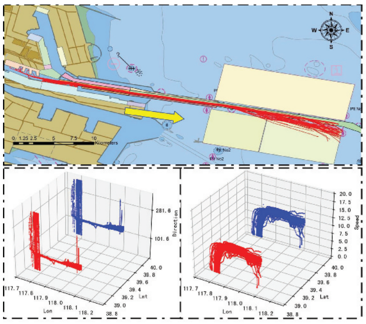

(40) Outbound pattern 7 of Container

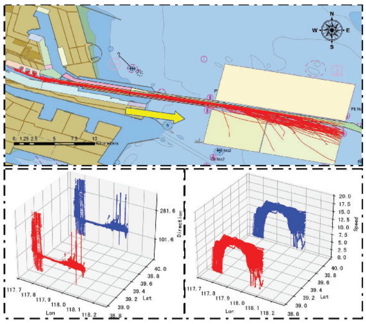
Figure A2c.

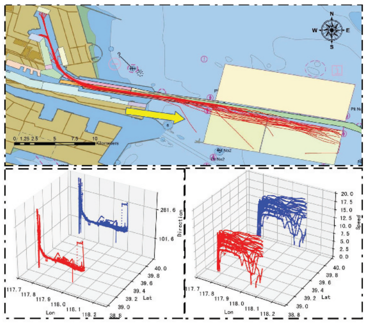


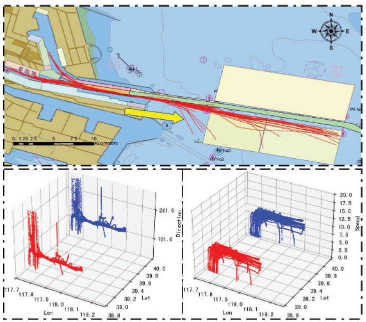
(1) Inbound pattern 1 of Cargo
(2) Inbound pattern 2 of Cargo
(3) Inbound pattern 3 of Cargo
(4) Inbound pattern 4 of Cargo
(5) Inbound pattern 5 of Cargo
(6) Inbound pattern 6 of Cargo
(7) Inbound pattern 7 of Cargo
(8) Inbound pattern 8 of Cargo
(9) Inbound pattern 9 of Cargo
(10) Inbound pattern 10 of Cargo
(11) Inbound pattern 11 of Cargo
(12) Inbound pattern 12 of Cargo
(13) Inbound pattern 13 of Cargo
(14) Inbound pattern 14 of Cargo
(15) Outbound pattern 1 of Cargo

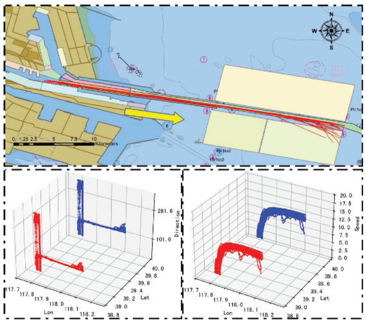
Figure A3a. The recognition results of ship traffic patterns based on the STPMTS method.

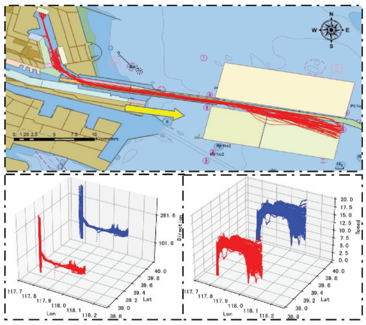

(16) Outbound pattern 2 of Cargo

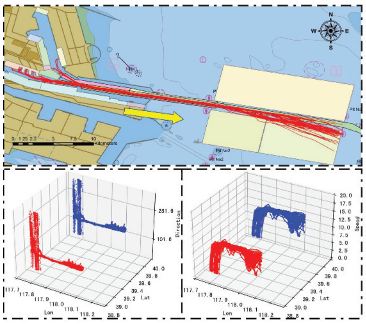

(17) Outbound pattern 3 of Cargo

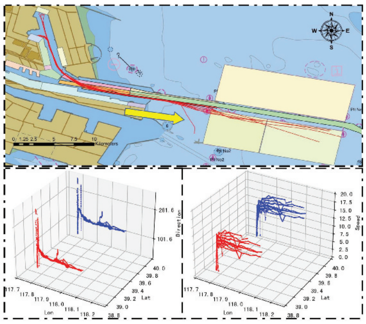

(18) Outbound pattern 4 of Cargo

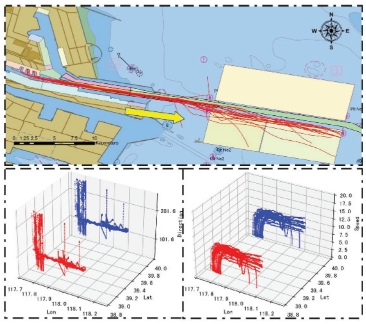

(19) Outbound pattern 5 of Cargo

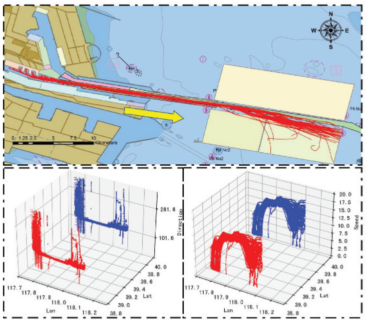

(20) Outbound pattern 6 of Cargo

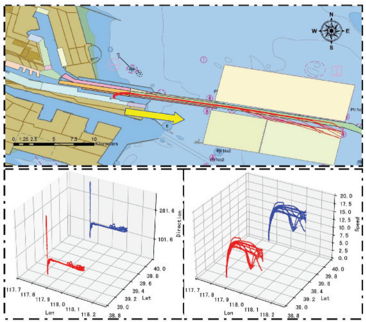

(21) Outbound pattern 7 of Cargo

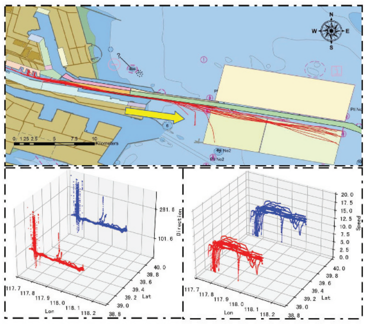

(22) Outbound pattern 8 of Cargo

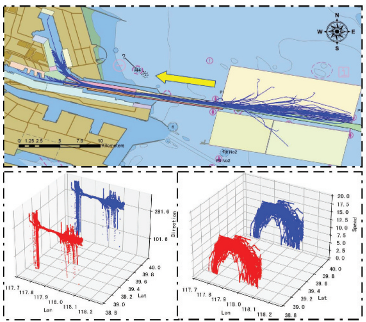

(23) Outbound pattern 9 of Cargo

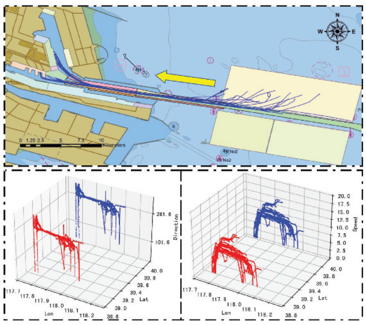

(24) Outbound pattern 10 of Cargo


(25) Outbound pattern 11 of Cargo


(26) Outbound pattern 12 of Cargo


(27) Outbound pattern 13 of Cargo


(28) Inbound pattern 1 of Container


(29) Inbound pattern 2 of Container

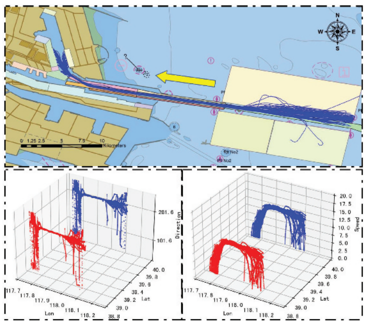
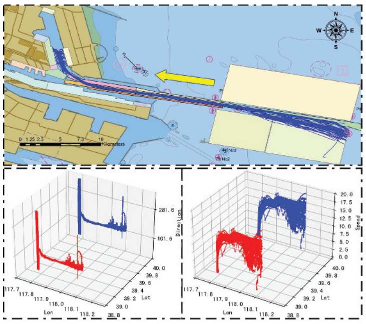
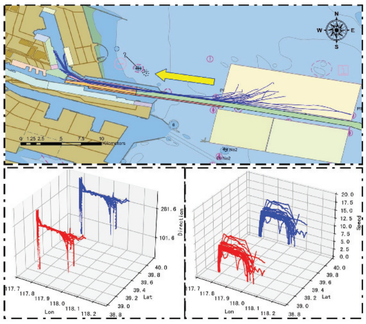

(30) Inbound pattern 3 of Container

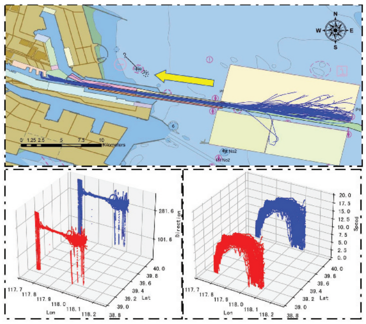
Figure A3b.



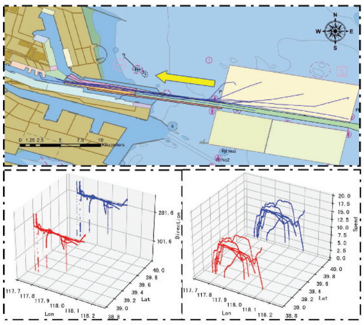
(31) Inbound pattern 4 of Container



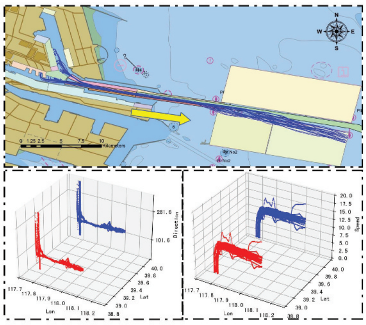
(32) Inbound pattern 5 of Container



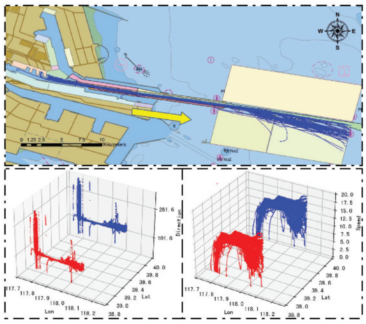
(33) Inbound pattern 6 of Container



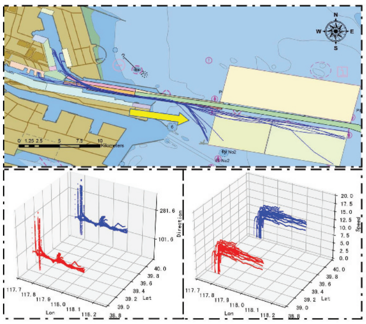
(34) Inbound pattern 7 of Container



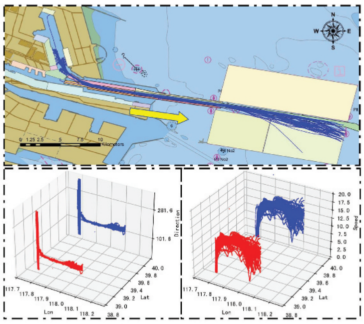
(35) Outbound pattern 1 of Container



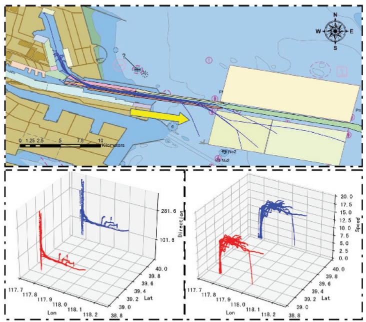
(36) Outbound pattern 2 of Container



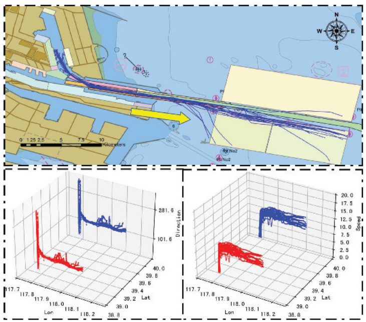
(37) Outbound pattern 3 of Container



(38) Outbound pattern 4 of Container



(39) Outbound pattern 5 of Container



(40) Outbound pattern 6 of Container

Figure A3c.