



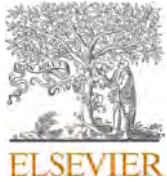
Shipping map: An innovative method in grid generation of global maritime network for automatic vessel route planning using AIS data

Downloaded from: <https://research.chalmers.se>, 2025-02-22 19:47 UTC

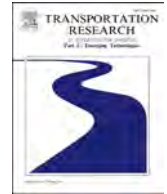
Citation for the original published paper (version of record):

Liu, L., Zhang, M., Liu, C. et al (2025). Shipping map: An innovative method in grid generation of global maritime network for automatic vessel route planning using AIS data. *Transportation Research, Part C: Emerging Technologies*, 171. <http://dx.doi.org/10.1016/j.trc.2025.105015>

N.B. When citing this work, cite the original published paper.

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Shipping map: An innovative method in grid generation of global maritime network for automatic vessel route planning using AIS data

Lei Liu^a, Mingyang Zhang^b, Cong Liu^{b,*}, Ran Yan^a, Xiao Lang^c, Helong Wang^d

^a School of Civil and Environmental Engineering Nanyang Technological University Singapore

^b Department of Mechanical Engineering Marine and Arctic Technology Group Aalto University Espoo Finland

^c Department of Mechanics and Maritime Sciences Chalmers University of Technology Gothenburg Sweden

^d NAPA Ltd Helsinki Finland

ARTICLE INFO

Keywords:

Maritime transportation
Maritime shipping network
Vessel routing planning
AIS data
CKBA-DBSCAN

ABSTRACT

Considering the challenges faced by current global grid-based route planning methods, including vessel navigability underestimation and high computational demands for fine grid configuration, this study introduces an innovative approach to the grid generation of a global maritime network for automatic vessel route planning. By leveraging global Automatic Identification System (AIS) data, the methodology focuses on advanced trajectory segmentation, waypoint detection, clustering algorithms, and route searching. A novel spatiotemporal approach is proposed to facilitate effective trajectory segmentation despite data discontinuities. The Pruned Exact Linear Time (PELT) algorithm is employed to identify waypoints, managing their quantity during heading instability. To recognize crucial berthing areas in ports and strategic waypoint zones at sea, a customized KNN-block adaptive Density-Based Spatial Clustering of Applications with Noise (CKBA-DBSCAN) is developed to address the challenges of varying density clustering parameters and high computational costs. Lastly, the double-layer network matching technique, which starts with grid-based route planning and refines to the final navigable and smoothed route, uniquely integrates data-driven and model-based strategies. Rigorous testing with a year's worth of global AIS data demonstrates high efficiency in planning navigable routes for various vessel types on worldwide voyages. The results underscore the practicality of the proposed approach in real-world route planning and maritime shipping network development. Remarkably, the methodology achieves a minimum 17.08 % reduction in time for global route generation. This hybrid approach, which integrates the strengths of both data-driven and model-based methods, significantly enhances vessel scheduling and routing efficiencies, showcasing its superior performance in comparative studies and its potential for widespread adoption in the maritime industry.

1. Introduction

The maritime transport sector plays a vital role in the global economy, facilitating the movement of goods and commodities across the world's oceans (UNCTD, 2022; Zhang et al., 2025). For the sector to succeed, efficient and safe shipping operations are essential,

* Corresponding author at: Otakaari 4, 02150, Koneteknikka 1, Espoo, Finland.

E-mail address: cong.l.liu@aalto.fi (C. Liu).

<https://doi.org/10.1016/j.trc.2025.105015>

Received 28 November 2023; Received in revised form 16 December 2024; Accepted 16 January 2025

Available online 30 January 2025

0968-090X/© 2025 The Author(s).

Published by Elsevier Ltd.

This is an open access article under the CC BY license

(<http://creativecommons.org/licenses/by/4.0/>).

Nomenclature

Variable Definition

AIS	Automatic Identification System
KNN	The k-nearest neighbours
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
CKBA-DBSCAN	Customized KNN-block adaptive DBSCAN
PELT	Pruned Exact Linear Time
A^*	A star
N_{stop}	The number threshold for the continuous berthing point
t_{stop}	The cumulative berthing duration
T_{stop}	The threshold for the continuous berthing duration
d_{stop}	The cumulative spatial distance
t_{aj}	The time interval between adjacent point
T_{conn}	The longest continuous time interval
d_{aj}	The spatial distance between adjacent point
D_{stop}	The spatial distance threshold between berthing points
D_{conn}	The maximum distance between consecutive adjacent points
S_{min}	The minimum sailing speed threshold
D_k	The spatial distance threshold of k_{th} neighbor
$minPts$	The minimum number that become core objects in DBSCAN
ϵ	The neighborhood radius threshold in DBSCAN
M	The clustering metric index
S_{bp}	The set of berthing points
$d_{i,i}$	The average distance from i_{th} point to other points within the same cluster
$d_{i,e}$	The average distance from i_{th} point to other points in other clusters
x	The proportion of anomalies to all data
α	The coefficient of proportion
D_{bel}	The distance threshold for a point to be considered within a berthing area
p_{center}^i	The centroid of the berthing area $area_{be}^i$
b_o^i	The boolean value representing whether point p_o is within $area_{be}^i$
T_k	The k_{th} trajectory
p_k^l	The l_{th} point in trajectory T_k
$NC_{p_k^l p_k^{l+1}}$	The navigation course from p_k^l to p_k^{l+1}
$lon_{p_k^l}$	The longitude of the l_{th} point in trajectory T_k
$lat_{p_k^l}$	The latitude of the l_{th} point in trajectory T_k
Y	The time sequence
$c(y_{t_j+1})$	The sum of fitting losses
$area_{wp}^n$	The n_{th} waypoint area
I	The for assessing waypoint detection performance
n	The number of detected waypoints
T_o	The original trajectory for waypoint detection
T_w	The waypoint-connected trajectory
$node^n$	The n_{th} node in network
$f(n)$	The comprehensive priority of node n
$g(n)$	The cost of node n from the starting node
$h(n)$	The estimated cost of node n from the ending node

with route planning emerging as a critical function (Tran et al., 2023; Liu et al., 2024). Effective route planning involves identifying the most suitable navigable routes between the designated berthing points (Dong et al., 2021; Du et al., 2022; Wang et al., 2022a). This process is instrumental in addressing several challenges that often arise during shipping operations, such as high fuel costs, carbon emissions, shipping delays, and security risks like piracy (Lehtola et al., 2019). Moreover, route planning can also aid in evaluating the accessibility and complexity of maritime shipping networks on a global scale (Xu et al., 2020), as shown in Fig. 1.

By identifying navigable and smoothed routes between regions, a comprehensive analysis of maritime shipping networks can be performed, enabling the identification of areas that require shipping network optimisation, aiming to improve the overall efficiency and functionality of maritime transportation (Liu et al., 2023c; Xu et al., 2020). Consequently, route planning can assist in improving the capacity and resilience of global maritime shipping networks, ensuring they can meet the growing demand for transportation

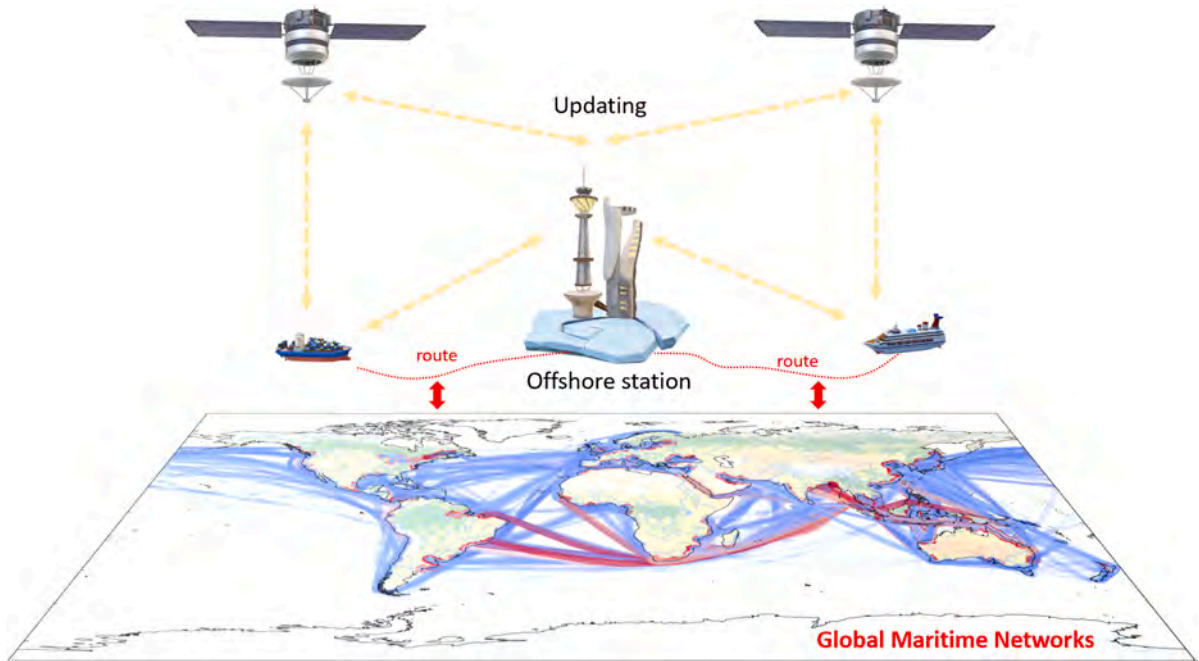


Fig. 1. The schematics of interactions between global vessel route planning and maritime shipping network (Container ships).

services (Alderson et al., 2019). Based on this context, the paper proposes a data-driven and model-based approach for extracting maritime shipping network data and automating global vessel route planning. This approach utilizes historical route records to extract maritime shipping network data and generate optimal routes. By leveraging data and models, the proposed approach aims to enhance the effectiveness of global maritime shipping networks, improve operational efficiency, and ensure safer and more reliable shipping operations (Dui et al., 2021; Jiang et al., 2021).

1.1. Literature review

Traditionally, vessel route planning has been a manual process conducted by the vessel's crew, relying on various documents and archival records of vessel routes. While these qualitative methods are deeply rooted in tradition, they exhibit notable limitations, particularly in regions with sparse historical navigational data or when detailed planning is necessary for complex port-to-port routes (Zhang et al., 2021a).

In contrast, the modern shipping industry increasingly adopts more advanced algorithms and intelligent navigation decision-making systems for weather routing and multi-objective route optimisation (Tsou & Hsueh, 2010; Wang et al., 2020a; Zhao et al., 2021; Zhang et al., 2024; He et al., 2024). Dynamic programming (DP), first proposed by Bellman (1952), exemplifies development and application in maritime contexts for these purposes (Meng & Wang, 2011; Papageorgiou et al., 2015; Wang et al., 2021). Additionally, intelligent navigation decision-making methods like machine learning (ML) and reinforcement learning (RL) offer dynamic adaptability and superior path rationality (Wang et al., 2020b). Chen et al. (2019) utilized tabular RL algorithms for real-time optimal route selection, adapting to dynamic maritime conditions. To address path planning in extensive state space scenarios, Li et al. (2021) and Gao et al. (2022) applied Deep Q Network (DQN) and Dueling DQN, respectively. Despite their advancements, RL algorithms require extensive training and computational resources, limiting their application to confined maritime areas. Moradi et al. (2022) designed the state as a multidimensional column vector to guide action exploration and reduce learning time, facilitating RL application in large-scale maritime environments. However, effectively applying algorithmic search-based methods or intelligent decision-making support methods across diverse and general scenarios remains a significant challenge.

Grid-based graph search algorithms form the cornerstone of algorithmic search-based methods and intelligent navigation decision-making support methods for weather routing and multi-objective route optimisation. These algorithms partition the maritime area into discrete segments and temporal stages, utilizing either a dynamic or a predefined static grid system. By integrating weather and environmental factors at the optimisation objective level, these algorithms offer robust solutions to route planning challenges. This study aims to advance these methods by focusing on the grid generation of a global maritime network for automatic vessel route generation. The following sections present a comprehensive review of existing grid and route generation methods, highlighting the importance of the proposed approach and its potential impact on the maritime industry.

Dynamic grid-based methods conduct route searches iteratively, updating based on the vessel's current position and navigating the vessel toward its destination. At each time stage, waypoints are generated from the current step; thus, the grid is updated iteratively, and the route is adjusted accordingly until the destination is reached. A notable example is the isochrone method, which involves

contours representing the furthest waypoints reachable in a defined time stage from different directions. It was initially introduced by James (1957) and later adapted by Hagiwara (1989). However, one drawback, known as the “isochrone loop”, occurs when the number of isochrones increases, potentially leading to impractical outcomes (Wang et al., 2022b; Poulsen et al., 2022). In recent years, more sophisticated methods have been developed for vessel routing. For instance, Zhang et al. (2021a) introduced an enhanced ant colony optimisation (ACO) algorithm to tackle the multi-objective vessel weather routing optimisation challenge, while Zhang et al. (2022) proposed a three-dimensional ACO algorithm for ice routing. Similarly, Xin et al. (2019) and Ma et al. (2021) have applied advanced genetic algorithms (GA), and Szlapczynska and Smierzchalski (2007), along with Lin et al. (2021), have incorporated multi-objective evolutionary algorithms (EA) in voyage planning. Additionally, to reduce fuel consumption while considering weather conditions, Wei et al. (2023) employed digital twin framework in decision-making, and both Chen et al. (2019) and Moradi et al. (2022) have utilised reinforcement learning techniques. These complex algorithms enable finer control over multiple parameters and facilitate frequent adjustments throughout the voyage planning process. However, the high computational demands and the necessity for continuous adjustments in navigational settings may render these advanced optimisation systems impractical for real-world maritime operations (Turna, 2023). Specifically, ocean-going vessels, which are generally large and heavily laden, may not be able to adapt quickly to the dynamic sailing conditions prescribed by these dynamic grid-based methods. This is because these methods underestimate vessel navigability. Additionally, frequent alterations in sailing status—such as speed, heading, and power—generated by these current algorithms necessitate continuous manoeuvring, leading to distorted trajectories resembling serpentine patterns, which in turn can increase fuel consumption, elevate emissions, and escalate operational risks.

Therefore, algorithms that offer high computational efficiency and simplified vessel navigation controls, such as predefined static grid-based methods, continue to be widely employed in voyage planning. Notably, two prominent algorithms that utilize static grid systems are Dijkstra’s algorithm (Dijkstra, 1959) and the A* algorithm (Hart et al., 1968), both of which have been effectively used to plan vessel routes. The Dijkstra algorithm, for example, has been adapted to accommodate variations in speed for weather routing (Wang et al., 2019) and to facilitate simultaneous route planning (Ma et al., 2020; Charalambopoulos et al., 2023). Park and Kim (2015) implemented the A* algorithm alongside a staged speed scheduling method to optimise fuel consumption while considering environmental conditions, which can be used for vessel route optimisation. Similarly, Bentin et al. (2016) applied the A* algorithm on a bulk carrier, effectively reducing fuel consumption by incorporating the wind-assisted propulsion system (WASP). Additional applications of the A* algorithm in route planning have been explored by Pennino et al. (2020), Shin et al. (2020), and Grifoll et al. (2022), Xu et al. (2024). These algorithms are advantageous for their ease of implementation in both single and multiple objective voyage optimisations. However, the effectiveness of the optimisation results and the computational load of static grid-based methods heavily depend on the grid configuration—specifically, factors such as the resolution of the grid, the number of nodes, and the extent of the search area covered by the grid. Based on the above, static grid-based methods face difficulties in generating smoothed vessel routes, often resulting in the vessel’s inability to accurately track the planned path at sea (Funk, 2017; Zhou et al., 2019).

Hence, in recent years, data-driven methods have emerged using AIS data that records both dynamic and static information on vessel navigation to construct potential navigable area and static grid system (Liu et al., 2023a; Wang et al., 2022b). It is implemented in two ways based on whether a maritime shipping network is constructed (Andersson & Ivehammar, 2017; Zhang et al., 2019; Zhang et al., 2018; Bläser et al., 2024). Without a maritime shipping network, navigable area and grid system can be performed by analysing vessel trajectories and recognising vessel states, dividing the operational areas into grids, and constructing a corresponding directed graph (Cai et al., 2021; Kaklis et al., 2024; Liu et al., 2023b; Naus, 2019; Li et al., 2022). However, this method encounters limitations in regions with complex navigation patterns. Alternative approaches, such as grid division, have addressed these complexities. Within the grid division method, navigable routes can be generated by detecting and recognising route points for vessel manipulation behaviour and then constructing grid system and maritime shipping networks. This method has been used by Varlamis et al. (2020) and Filipiak et al. (2020) in local waters and by Yan et al. (2020), Liu et al. (2023b), and Bläser et al. (2024) in global open water scenarios. However, most of these studies lack a detailed description of the method for generating the connection attributes of static grid system, making it challenging to apply to navigable routes between all areas from a global perspective.

Current, the main methods for identifying route waypoints are threshold detection (Zhao & Shi, 2019) and trajectory compression (Yan et al., 2022; Bläser et al., 2024; Ji et al., 2022; Truong et al., 2020). The application of threshold detection in maritime navigation faces significant challenges. Firstly, it is challenging to establish a unified parameter for identifying route points across scenarios with multiple vessels. This complexity arises due to the variability in vessel movements and patterns. Secondly, there is often an over-detection of waypoints in areas of high vessel convergence, such as near ports or busy shipping lanes. This leads to an excessive number of waypoints being identified, which can complicate navigation and route planning. Additionally, the issue of a non-uniform threshold is also present in trajectory compression algorithms, as highlighted by Tang et al. (2021). This inconsistency in the threshold makes it difficult to apply a standardized approach across different maritime scenarios. For constructing waypoint areas, density clustering methods like DBSCAN and Ordering Points to Identify the Clustering Structure (OPTICS) are commonly used. However, these methods struggle in open water areas where the distribution density of waypoints varies significantly. This variation highlights the unreliability of networks constructed using unified parameters and underscores the need for a parameter-adaptive clustering to ensure accurate identification of nodes in the networks (Bläser et al. 2024).

1.2. Research gaps and contributions

Several research gaps remain to be addressed in the area of predefined static grid-based route planning methods. Current models often lack the flexibility and scalability necessary for application across various sailing regions and scenarios, primarily due to the underestimation of vessel navigability. These methods require significant computational resources, as optimised routes depend heavily

on grid configuration. Achieving optimal routes necessitates a high resolution of the grid, a large number of nodes, and extensive coverage of the search area, complicating real-time route planning, particularly in large-scale shipping operations.

Implementing data-driven methods to construct potential navigable areas and static grid systems introduces additional challenges. For instance, identifying route waypoints and building maritime shipping networks for global navigable area identification remain unresolved. Most current methods focus predominantly on local or regional waters, leaving a noticeable gap in research on global route planning, which is crucial for extensive maritime operations. Moreover, there is a lack of consensus on generating connections between nodes, such as berthing areas at ports and waypoint areas at sea, in the construction of maritime shipping networks for grid generation in route planning. Existing methods for identifying route waypoints also suffer from parameter standardisation issues, making application across different areas challenging.

Therefore, there is an urgent need for more comprehensive evaluation metrics to assess the effectiveness and reliability of globally generated vessel routes. These metrics would help refine existing methodologies and ensure they meet the operational demands of the modern maritime shipping industry.

This study proposes an innovative approach to grid generation of global maritime network for automatic vessel route planning utilising AIS data. This approach synergistically combines data-driven methods for navigable area grid generation and static grid-based route planning method to develop highly efficient and reliable global route planning systems, specifically for large-scale shipping operations. The proposed method addresses several shortcomings of existing research in this field and the specific contributions are summarized in the following three aspects:

- Introduces a novel approach to grid generation of global maritime network for automatic vessel route planning by leveraging AIS data. The approach combines data-driven methods for generating navigable area grids and static grid-based route planning, enhancing the flexibility, scalability, and efficiency of global maritime operations. This method addresses the previous research gap concerning the lack of comprehensive global route planning systems.
- Presents a new trajectory segmentation method that efficiently divides berthing and non-berthing segments based on spatial and temporal distances between waypoints. Additionally, it introduces a waypoint identification method using CKBA-DBSCAN clustering algorithm. These methods surpass the limitations of existing models in identifying and constructing navigable waypoints, significantly improving the accuracy and applicability of route planning.
- Extends the search space for navigable routes through the integration of maritime shipping networks with grid networks, using the A* algorithm. This not only addresses the challenges of limited network accessibility but also enhances the reliability of the route planning system. This contribution specifically tackles the previously unaddressed issue of effectively connecting nodes like berthing areas and waypoints in maritime shipping networks.

The proposed methods effectively address the challenges of global route planning by providing a reliable and efficient method for identifying navigable waypoints and constructing maritime shipping networks. By combining data-driven methods for constructing potential navigable areas and static grid systems and static grid-based A* algorithm for route planning, the proposed approach provides a comprehensive and effective solution to global vessel route planning, which can significantly improve the efficiency, safety, and cost-effectiveness of shipping and has the potential to be applied in practical maritime transportation systems.

2. Methodology

The framework of the proposed method is shown in Fig. 2. The steps of implementing the data-driven and model-based vessel route planning method are summarized as follows:

Step I: AIS data preprocessing and vessel trajectory segmentation

Historical voyage records are acquired from AIS by identifying the berthing points. This study proposes a spatiotemporal methodology to achieve this on a global scale. This method facilitates the identification of berthing points through a combination of spatial and temporal distance analysis, enabling effective trajectory segmentation in multi-class scenarios. Once the berthing points have been identified, it becomes feasible to extract global trajectory segments. These segments can then be utilised to develop a global vessel route matching model and maritime shipping network.

Step II: Data-driven global vessel route matching method

This method utilises the berthing points information of each segment to identify berthing areas through a CKBA-DBSCAN clustering algorithm, which demonstrates better applicability than the traditional DBSCAN clustering in scenarios with significant variations in the density of points across different regions in large waters. Building upon this, an effective route matching method is also utilised to extract historical vessel trajectories with the same origin and destination points.

Step III: Model-based global vessel route generation

A model-driven approach is used to generate a global vessel route. To address the challenge of standardising parameters for heading threshold detection and trajectory compression detection of waypoints in large water bodies, change point detection is employed to identify waypoints. Based on this, berthing areas and waypoint areas are constructed as nodes and connected by navigable routes based on historical sailing trajectories. Those nodes and routes form the global maritime shipping network, and the A* algorithm is employed to generate vessel routes on this network.

Step IV: Data-driven and model-based vessel route planning method

A vessel route planning method is developed by adapting both data-driven and model-based approaches. Specifically, a global grid network method is employed to integrate these two approaches, resulting in the proposed double-layer network method for vessel

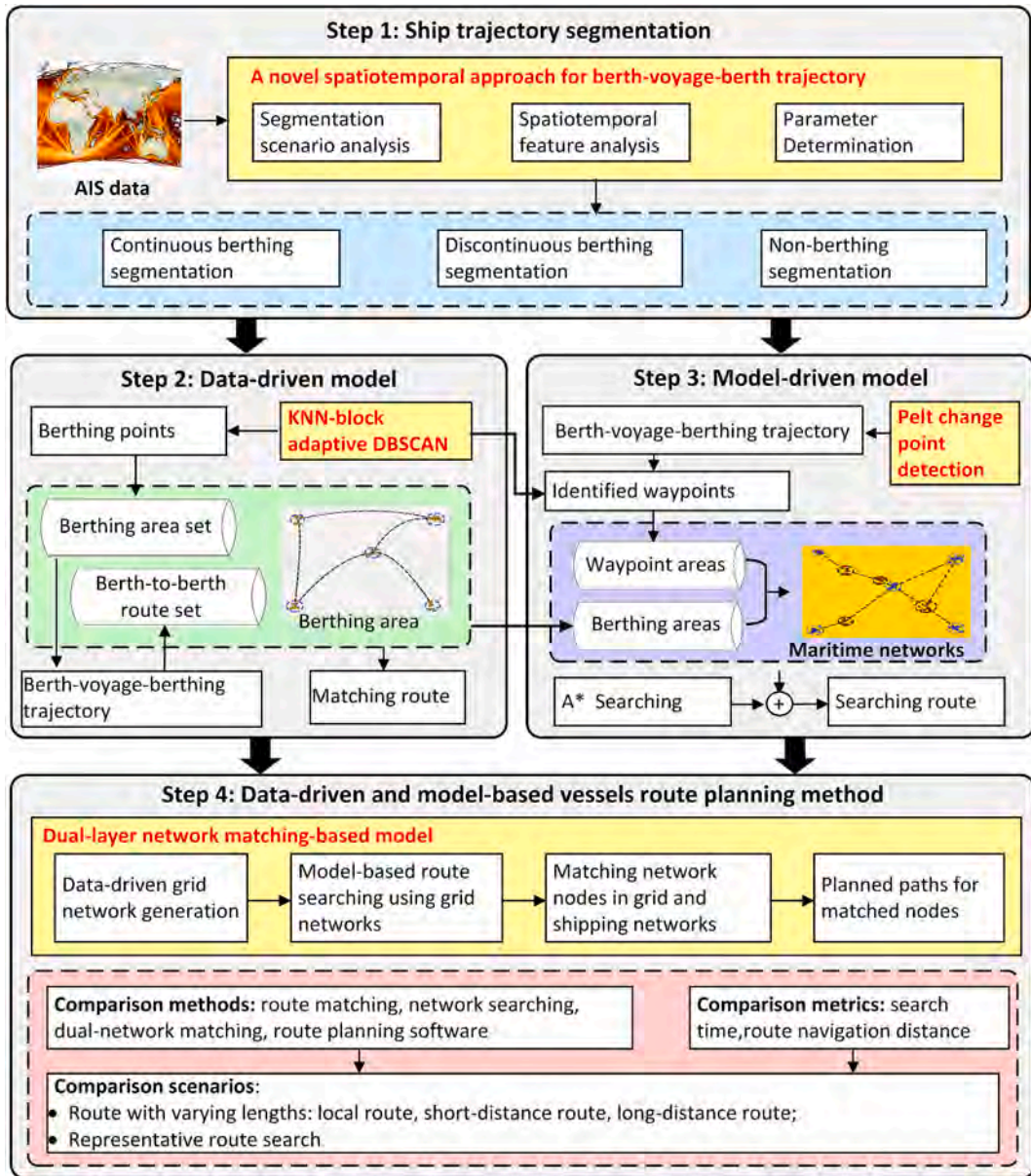


Fig. 2. The flowchart of data-driven and model-based methods for automatic global vessel route planning with AIS data.

route planning, and enhancing the probability of route searching. To evaluate the performance of the proposed method, a comparative study between the data-driven, model-based, and double-layer networks is conducted under various scenarios.

2.1. Vessel trajectory segmentation

Vessel trajectory segmentation is a critical step in analysing and understanding vessel moving voyages, which can help the development of an end-to-end navigable route database and the identification of network nodes. In the following sections, we introduce the schematic of vessel trajectory segmentation using AIS and a novel segmentation method accounting for spatiotemporal features.

2.1.1. Scenario analysis of vessel trajectory segmentation

Vessel trajectory segmentation involves dividing a complete vessel trajectory into multiple continuous segments, with each segment representing the “berth-voyage-berth” pattern of vessel movement. Therefore, by identifying AIS data points corresponding to the berthing behaviour within the trajectory, the segmentation process can be effectively executed to obtain the desired results. One

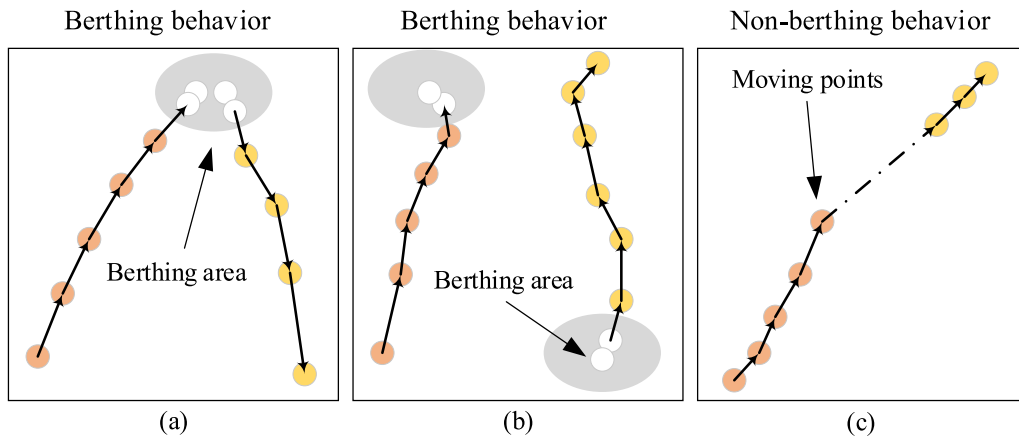


Fig. 3. Vessel trajectory segmentation encompasses three distinct scenarios: (a) spatially continuous berthing segmentation, (b) spatially discontinuous berthing segmentation, and (c) non-berthing segmentation.

Table 1
Characteristics of Berth/ Non-Berthing Segmentation Points.

Segmentation type	Temporal features	Spatial features	Voyage information
Spatially Continuous Berthing Segmentation	<ol style="list-style-type: none"> 1. Total large time interval. 2. Spatially continuous berthing segments have small time intervals between adjacent points. 3. Spatially continuous berthing segments have larger time intervals between non-adjacent points. 	Spatial position remains relatively unchanged.	Consistent or inconsistent destination.
Spatially Discontinuous Berthing Segmentation	<ol style="list-style-type: none"> 1. Non-continuous time intervals. 2. Small time intervals between other berthing points. 	Large spatial distance between adjacent berthing points.	Different destinations.
Non-Berthing Segmentation	Non-continuous time intervals.	Large spatial distance between adjacent non-berthing points.	Consistent or inconsistent destination.

widely used method for vessel trajectory segmentation relies on the time interval of vessel points using AIS dynamic information. However, in practical applications, other factors may also influence the trajectory segmentation results. Fig. 3 illustrates two types of segmentation (three scenarios). For example, considerable time intervals or significant spatial distances between adjacent points can hinder the identification of intermediate voyage processes, leading to non-berthing segmentation. Consequently, the final segmented trajectory should consist of two distinct types of berthing segmentation resulting from berthing behaviour and non-berthing segmentation resulting from moving behaviour with long periods of missing data.

To distinguish between different types of trajectory segmentation, Table 1 summarizes the characteristics of segmentation points in various scenarios from AIS data. Berthing segmentation requires analysing the spatiotemporal characteristics of consecutive AIS points. Specifically, if the starting position of the k -th trajectory segment is within the same area as the ending position of the $(k-1)$ -th segment, it is classified as a spatially continuous berthing segmentation, as depicted in Fig. 3(a). Otherwise, it is considered a spatially discontinuous berthing segmentation, as shown in Fig. 3(b). In the latter scenario, a considerable spatial and temporal distance exists between the two adjacent points, indicating that they correspond to different journeys with likely different destinations (Yin et al., 2022; Zhang et al., 2020). For non-berthing segmentation, as illustrated in Fig. 3(c), if the last point of k -th trajectory segment and the first point of the $(k+1)$ -th segment are both in a navigation state, and there is a long temporal gap and significant spatial distance between them, these two segments are considered parts of the same journey if the destination remains unchanged. Conversely, if the destination differs, it implies missing data between the two separate journeys (Zhang et al., 2021b).

In the case of berthing segmentation, the number of consecutive AIS points, combined with their spatiotemporal distances, serves as a key characteristic to determine the presence of a berthing state. Whether a trajectory contains points in a berthing state depends on whether the vessel has turned off its AIS equipment after entering the port area. If the vessel continues to transmit data while in a berthing state, the trajectory that containing a large number of data points will mostly be in the berthing state, with the location remaining relatively unchanged and the time intervals between points being relatively uniform. Conversely, if the vessel turns off the AIS equipment immediately after entering the port area, there will be no trajectory points in the berthing state. A considerable time interval will exist between the two points corresponding to turning on and off the AIS equipment.

2.1.2. Vessel trajectory segmentation based on spatiotemporal features

Based on the above analysis, this study proposes a trajectory segmentation method based on the temporal and spatial distance between AIS points, building upon the vessels berthing identification method proposed by (Huang et al., 2021). They determined the berthing points for trajectory segmentation by setting conditions for both the minimum number of berthing points and the minimum

berthing duration, which, however, cannot be applied to the situations where the number of berthing points is limited due turning off the AIS equipment immediately after vessel berthing. Additionally, in scenarios where voyage data is missing between two vessel points (under berthing or moving status), trajectory segmentation cannot be achieved. Therefore, their method is only applicable to berthing identification in scenarios with continuous AIS data in local area and cannot be used in scenarios where continuous AIS data cannot be guaranteed in larger-scale waters.

Algorithm 1 summarises the process of the proposed method, and the rules are specified as follows. (a) If the number of consecutive berthing points is no less than the minimum number N_{stop} of consecutive berthing points required to define a berthing phase, the segment is classified as spatially continuous berthing segmentation. (b) If the cumulative berthing duration t_{stop} of consecutive berthing points exceeds the minimum duration T_{stop} to define a berthing phase, and the cumulative spatial distance d_{stop} does not exceed the distance travelled at the minimum sailing speed S_{min} during t_{stop} , the segment is categorised as a spatially continuous berthing segmentation. Conversely, if d_{stop} exceeds this navigation distance, the segment is considered a spatially discontinuous berthing segmentation. (c) If the time interval t_{aj} adjacent points with exceeds the maximum allowable continuous time interval T_{conn} , and the spatial distance d_{aj} exceeds the distance travelled at S_{min} during t_{aj} , it is classified as a non-berthing segmentation. (d) All other scenarios are not considered for trajectory segmentation and are treated as part of a continuous trajectory. For instance, if the time interval t_{aj} between adjacent points exceeds T_{conn} but does not exceed T_{stop} , and the spatial distance d_{aj} does not exceed the distance travelled at S_{min} during that time period, it indicates a brief berthing state and is considered as a part of a single segment.

There are four parameters to be determined in the abovementioned rules, namely N_{stop} , T_{stop} , S_{min} , and T_{conn} . Notably, a point is considered a berthing point relative to the previous one if the distance between them does not exceed D_{stop} rather than its status data. Additionally, the maximum allowable continuous distance D_{conn} can be derived more easily from historical data than T_{conn} . Once D_{conn} is determined, T_{conn} can be calculated as $T_{conn} = D_{conn}/S_{min}$. Consequently, the five parameters to be determined are N_{stop} , D_{stop} , T_{stop} , S_{min} and D_{conn} . These parameters can be estimated through statistical analysis of AIS data, as detailed in Section 3.1.

2.2. Data-driven global vessel route matching method

For vessel route planning, when provided with a set of origin and destination, the simplest approach to search historical navigation trajectories is by employing a route search algorithm that can identify navigable routes between these specified locations. To achieve this, the given origin and destination should be spatially classified to determine their respective regions. Once the classification is completed, the algorithm can then identify the navigable routes between the corresponding berthing areas. As described in Section 2.1, all berthing points are already obtained and can be clustered to obtain the required berthing areas for this specific purpose.

Algorithm 1: Trajectory Segmentation Method Based on Spatial and Temporal Distance of Vessel Points

Input: AIS data, D_{stop} , D_{conn} , S_{min} , N_{stop} , T_{stop}
Output: Trajectory segment set U_{is} .

- 1: Generate an empty trajectory data U_{is}
- 2: Calculate the maximum allowable continuous time interval $T_{conn} = D_{conn}/S_{min}$
- 3: **for** AIS data T^i of vessel i **do**
- 4: Generate an empty list of segmentation points L_{sepe} and an empty temporary list of segmentation points L_{sus_sepe}
- 5: Add the first point p^1 in T^i to L_{sus_sepe} and initialize p_{stop} to p^1 .
- 6: **for** the second point p^j to the last point in T^i **do**
- 7: Calculate the spatial distance d_{stop} and time interval t_{stop} between p^j and p_{stop}
- 8: Calculate the spatial distance d_{aj} and time interval t_{aj} between p^j and p^{j-1}
- 9: **if** $d_{aj} < D_{stop}$ **then**
- 10: Add p^j to L_{sus_sepe}
- 11: **continue**
- 12: Return the number l_{stop} of points in L_{sus_sepe}
- 13: **if** $l_{stop} \geq N_{stop}$ **or** $t_{stop} \geq T_{stop}$ **then**
- 14: **if** $d_{stop} \leq S_{min} \cdot t_{stop}$ **then**
- 15: Mark the points in L_{sus_sepe} as spatially continuous berthing segmentation points
- 16: **else**
- 17: Mark the points in L_{sus_sepe} as spatially discontinuous berthing segmentation points
- 18: **else if** $t_{aj} \geq T_{conn}$ **and** $d_{aj} \geq S_{min} \cdot t_{aj}$ **then**
- 19: Mark the points in L_{sus_sepe} as non-berthing segmentation points
- 20: Add the points in L_{sus_sepe} to L_{sepe}
- 21: Clear L_{sus_sepe}
- 22: Set p^j as p_{stop}
- 23: Divide the trajectory T^i according to L_{sepe} to obtain the segmented trajectory U_{is}^i
- 24: Add U_{is}^i to U_{is}
- 25: Clear L_{sepe}
- 26: Return U_{is}

2.2.1. Vessel berthing point clustering

In the context of point clustering-based berthing area identification, where the number of berthing areas cannot be determined in advance due to variability and temporary berthing, density clustering is chosen as the preferred method. However, global water areas

may exhibit variations in berthing point density, leading to inconsistent clustering parameters for density clustering. Additionally, density clustering incurs higher computational costs, making it challenging to process all points simultaneously, especially when numerous berthing points are involved. To address these issues, it is necessary to partition the research water area and implement an adaptive hyperparameter selection strategy (Xin et al., 2024).

In response to these challenges, this study proposes a CKBA-DBSCAN clustering approach. Given that berthing points are typically concentrated, a KNN filtering method is first applied. A berthing point is considered valid and contributes to berthing area identification only if its distance to its k -th neighbor is less than a predefined threshold D_k . Subsequently, the entire water area is then divided into grids, with activated grids defined as those containing at least one valid berthing point identified through the KNN filtering process. The depth-first search is then employed to group adjacent activated grids into blocks. Following this, adaptive DBSCAN clustering is applied on each block for berthing area identification.

DBSCAN clustering involves the neighbourhood radius threshold ε and the minimum number of points $minPts$. Although these two parameters are coupled and usually adjusted together to find the appropriate cluster structure, adaptively determining both parameters simultaneously need high computational effort. Therefore, the value of $minPts$ is initially determined based on existing literature (Li et al., 2021), focusing on ensuring a sufficient density of points to form a valid berthing area. To make ε adaptive for each block, an improved Silhouette score M is proposed to determine the optimal value for each block, as expressed in Equation (1).

$$M = \frac{\sum_{i \in S_{bp}} ((d_{i,e} - d_{i,i}) / \max\{d_{i,i}, d_{i,e}\})}{|S_{bp}|} - e^{\alpha(x-1)} \quad (1)$$

where $d_{i,i}$ and $d_{i,e}$ donate the average distance from the i -th sample to all other points within the same cluster and the average distance from the i -th sample to all points in other clusters, respectively. S_{bp} represent the set of berthing points, with $|S_{bp}|$ indicating the set length. x represents the proportion of anomalies to all data, and α represents the coefficient of proportion. The first component of Equation (1) is derived from the Silhouette score. This choice reflects the observation that normal berthing points should be distributed across different port anchorages, with distances between points in the same category significantly smaller than those between points in different categories. The second component introduces a penalty term designed to reduce the likelihood of erroneously classifying a substantial portion of points as anomalies when relying solely on the Silhouette score. By defining a specific range for the ε values and performing clustering with varying ε values within this range, the clustering results are evaluated using Equation (1). The optimal ε is then determined based on the clustering metric index results.

After completing the clustering for berthing points, the centre of each cluster is taken as the centre of the corresponding berthing area, while the maximum distance from the centre to any points within the cluster is calculated as the radius of the berthing area. Finally, all berthing areas $\{area_{bp}^1, area_{bp}^2, \dots, area_{bp}^n\}$ are obtained.

2.2.2. Navigable routes matching

Once the berthing areas are identified, the original trajectories are categorised based on the berthing areas where their starting and ending points are located. Using this information, a navigable route database is constructed to store all historical trajectories between pairs of berthing areas. For the intended search route's origin and destination, the corresponding berthing areas are determined by calculating their proximity to the identified berthing areas. The navigable routes between these two berthing areas are then retrieved from the navigable route database. If multiple historical trajectories exist between the same pair of berthing areas, the selection of the navigable route can be based on predefined criteria, such as the shortest distance or the highest frequency of use.

2.3. Model-driven global vessel route generation method

The effectiveness of route planning based on trajectory matching depends on the presence of navigation trajectories between the target areas. However, in cases where historical sailing trajectories are unavailable, navigable routes can still be generated by identifying waypoints within the target water area and considering the connections between these points through the extraction of a global maritime shipping network.

2.3.1. Waypoint identification based on change point detection

Navigation areas refer to specific areas of the ocean or waterways that are designated for shipping routes. To identify navigation areas, it is essential to locate navigation points from historical sailing data to construct these areas. This study introduces a waypoint detection approach aimed at identifying navigation points.

Before detecting waypoints, the course sequence of a trajectory needs to be obtained first. Although AIS data provides the vessel course over ground, this value represents only the vessel's instantaneous state. Given the k -th trajectory data $T_k = (p_k^1, p_k^2, \dots, p_k^l)$, the course from point p_k^l to p_k^{l+1} is calculated using Equation (2) (Yan et al., 2020).

$$NC_{p_k^l, p_k^{l+1}} = \arctan \frac{\sin(x_{p_k^{l+1}} - x_{p_k^l})}{\cos(y_{p_k^l}) \times \tan(y_{p_k^{l+1}}) - \sin(y_{p_k^l}) \times \cos(x_{p_k^{l+1}} - x_{p_k^l})} \quad (2)$$

where $(x_{p_k^l}, y_{p_k^l})$ and $(x_{p_k^{l+1}}, y_{p_k^{l+1}})$ represent the location of points p_k^l and p_k^{l+1} , respectively. This yields the corresponding navigation

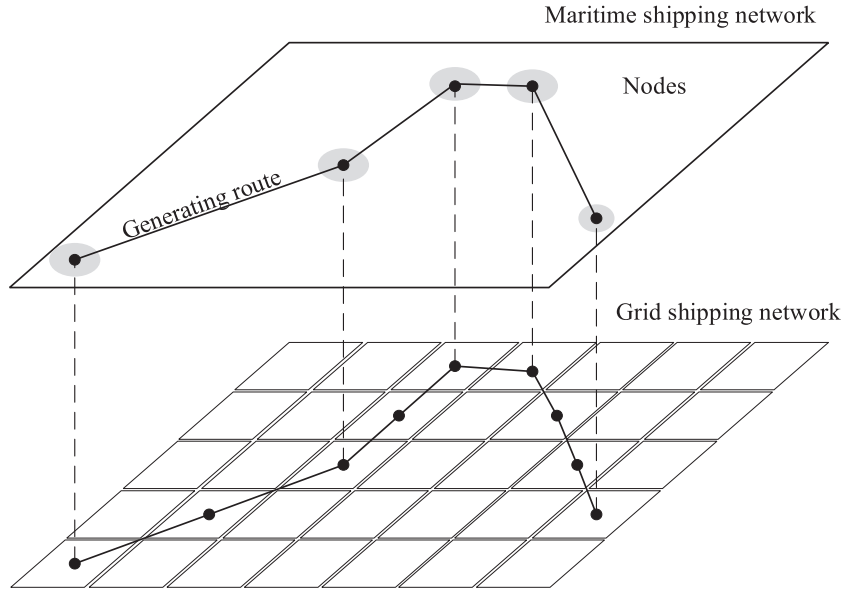


Fig. 4. Schematic diagram of double-layer network matching for searching navigable paths.

course sequence $(NC_{p_k^1 p_k^2}, \dots, NC_{p_k^l p_k^{l+1}}, \dots, NC_{p_k^{l-1} p_k^l})$.

Building on this, the Pruned Exact Linear Time (PELT) algorithm, a robust change point detection method, is employed to identify waypoints. Change point detection can be formulated as finding a feasible partition Λ for sequence Y that minimizes the sum of fitting losses $c(y_{t_j t_{j+1}})$ for each time segment and a penalty factor $pen(\Lambda)$ that considers the complexity of the partition, as expressed in Equation (3).

$$\min_{\Lambda} \left\{ \sum_{j=0}^J c(y_{t_j t_{j+1}}) + pen(\Lambda) \right\} \quad (3)$$

With this concept, the PELT algorithm (Killick et al., 2012) is utilised to optimally detect the unknown number of change points, making it suitable for large-scale vessel trajectories of varying lengths and shapes.

Finally, a metric I is introduced to evaluate the performance of waypoint detection and facilitate comparisons across different methods, as defined in Equation (4).

$$I = n^* \log(\text{Hausdorff}(T_o, T_w)) \quad (4)$$

where n represents the number of waypoints, t_o and t_w denote the original trajectory and the waypoint-connected trajectory, respectively. The Hausdorff() distance quantifies the similarity between the two trajectories. This metric is designed to minimise the number of waypoints while maintaining a high degree of similarity to the original trajectory. A lower value of I indicates better performance in terms of both waypoint detection efficiency and trajectory fidelity.

2.3.2. Global maritime shipping network generation

The waypoints are grouped using CKBA-DBSCAN clustering to identify the waypoint areas $\{area_{wp}^1, area_{wp}^2, \dots, area_{wp}^n\}$. Notably, unlike berthing point processing, waypoint clustering considers both the spatial distance and the direction feature (Bläser et al., 2024). Suppose we have the waypoint sequence $(w_k^1, \dots, w_k^j, \dots, w_k^l)$ of the k -th trajectory, Equation (2) is applied to calculate the course values, with the course set to 0 for the initial waypoint w_k^1 . Based on this, the directional characteristics consist of two components: the navigation course $NC_{w_k^{j-1}, w_k^j}$ from w_k^{j-1} to w_k^j and the navigation course $NC_{w_k^j, w_k^{j+1}}$ from w_k^j to w_k^{j+1} . Therefore, a comprehensive distance is calculated by integrating spatial dissimilarity and course dissimilarity among the waypoints. The waypoint areas and berthing areas are then combined to form the set of nodes $\{node^1, node^2, \dots, node^n\}$ in the shipping network. The arcs between network nodes are extracted from the trajectory data. Specifically, when a trajectory departs from the node $node^i$ and reaches the node $node^j$, it indicates the existence of a connecting arc between $node^i$ and $node^j$. If multiple connecting arcs exist between the same pair of nodes, the shortest trajectory is selected as the connecting arc. Finally, after determining the sets of nodes and the corresponding connecting arcs, a directed graph of the shipping network is constructed.

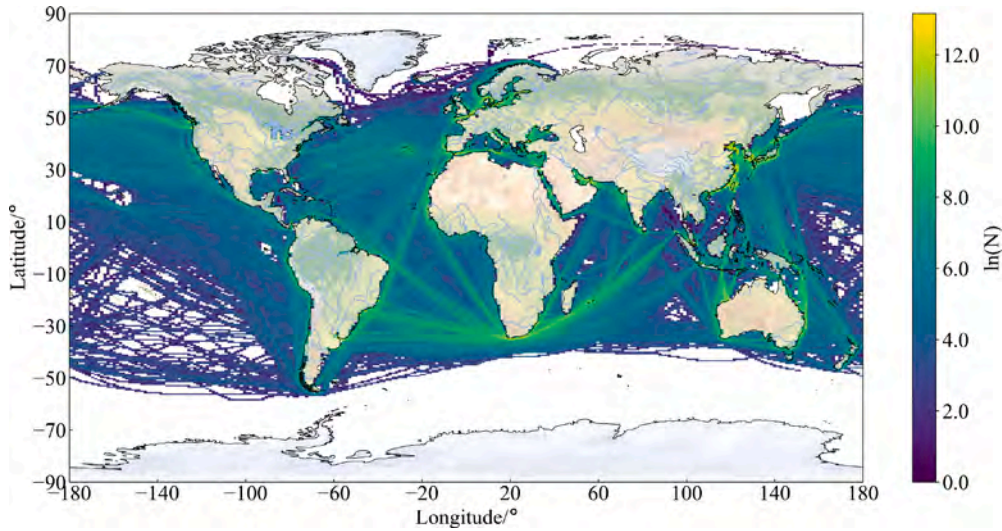


Fig. 5. Distribution of global bulk carrier and dry cargo vessel AIS data in 2018. (Data Source: seasearcher)

2.3.3. Vessel route generation using A* algorithm

Given the search origin and destination, and the determination of their corresponding nodes, we adopt the A* algorithm for route planning in the constructed shipping network. The A* algorithm utilises a heuristic function to guide the search towards the goal state, which leads to faster convergence to the optimal solution, as shown in Equation (5).

$$f(n) = g(n) + h(n) \quad (5)$$

where $g(n)$ represents the actual cost from the starting point to node n , which is obtained by summing the losses on the arcs from the starting point to node n as $\sum_{i=1}^n g_{i(i+1)}$. The loss on an arc from node i to node $i+1$ is denoted as $g_{i(i+1)}$. $h(n)$ represents the estimated cost from node n to the node where the ending point is located. Consequently, $f(n)$ denotes the total estimated cost of node n .

2.4. Data-driven and model-based vessel route planning method

Based on the shipping network described above, searching for navigable routes ensures the discovery of paths within the existing navigation network. However, the network's accessibility is limited since node connections in the shipping network are derived solely from trajectories. This can result in some unreasonable connections in the generated route. To address this, this study further proposes a navigable route planning approach based on a double-layer network matching method, as depicted in Fig. 4. The fundamental concept is to create routes using a grid network, and if a grid node is not a waypoint area, the vessel does not need to manoeuvre in that specific grid area. The trajectory generated by the grid network connects the centres of each grid, and reconnecting the waypoints avoids the requirement of smoothing the trajectory. By searching for navigable routes based on the grid network, the accessibility of the shipping network is expanded while minimising route loss to a greater extent.

The specific steps of the navigable route planning based on the double-layer network matching are as follows. To ensure better alignment between the two networks, the water area is first divided into grids identical to Section 2.2.1, with each grid treated as a node. Following this, the connections between grid nodes are constructed using trajectories to generate a grid network. After determined the grids that origin and destination belong to, the navigable route planning is performed using the grid network and the A* algorithm. The resulting path $\{node^1, node^2, \dots, node^k\}$ from the grid network is then mapped to the shipping network to form a new node set $\{node^{1'}, node^{2'}, \dots, node^{k'}\}$. If the origin or destination cannot be matched to a node in the shipping network, the corresponding labels from the grid network is added to the new node set, and the final route is generated using this updated node set.

3. Case studies and discussions

The experimental analysis uses the 2018 global dry bulk carrier AIS data as a case study, comprising approximately 50 million AIS data points, as illustrated in Fig. 5. Two key factors motivate the focus on dry bulk carriers as the main research subject. Firstly, in contrast to container vessels, dry bulk carriers predominantly engage in transporting goods directly from the origin port to the destination port. This operational characteristic makes the route planning for dry bulk carriers more straightforward, focusing primarily on point-to-point transportation without the complexities of multiple stops, significantly simplifying identifying direct navigation routes while searching for feasible paths. Secondly, dry bulk carriers transport goods to diverse regions worldwide, thereby expanding the scope of route planning possibilities.

Table 2
Dataset description after each processing step.

Processing step	Dataset content	Data content	Data volume
1 /	Raw AIS data	[IMO, date and time, location, SOG, COG, navigation status]	48,229,527
2 Trajectory segmentation	Segmented trajectories	[IMO, date and time, location, SOG, ..., trajectory label]	45,804,039
3 Berthing area recognition	Valid berthing points	[location, grid label]	6,949,260
	Berthing areas	[area label, center location, radian]	2,221
	Trajectories among areas	[IMO, ..., trajectory label]	19,184,389
4 Waypoint area recognition	Segmented trajectories for waypoint identification	[IMO, ..., trajectory label]	40,329,976
	Valid waypoints	[location, grid label]	373,243
	Waypoint areas	[area label, center location, radian]	2,588
5 Arc extraction	Arcs	[start area, end area, arc trajectory information]	286,707

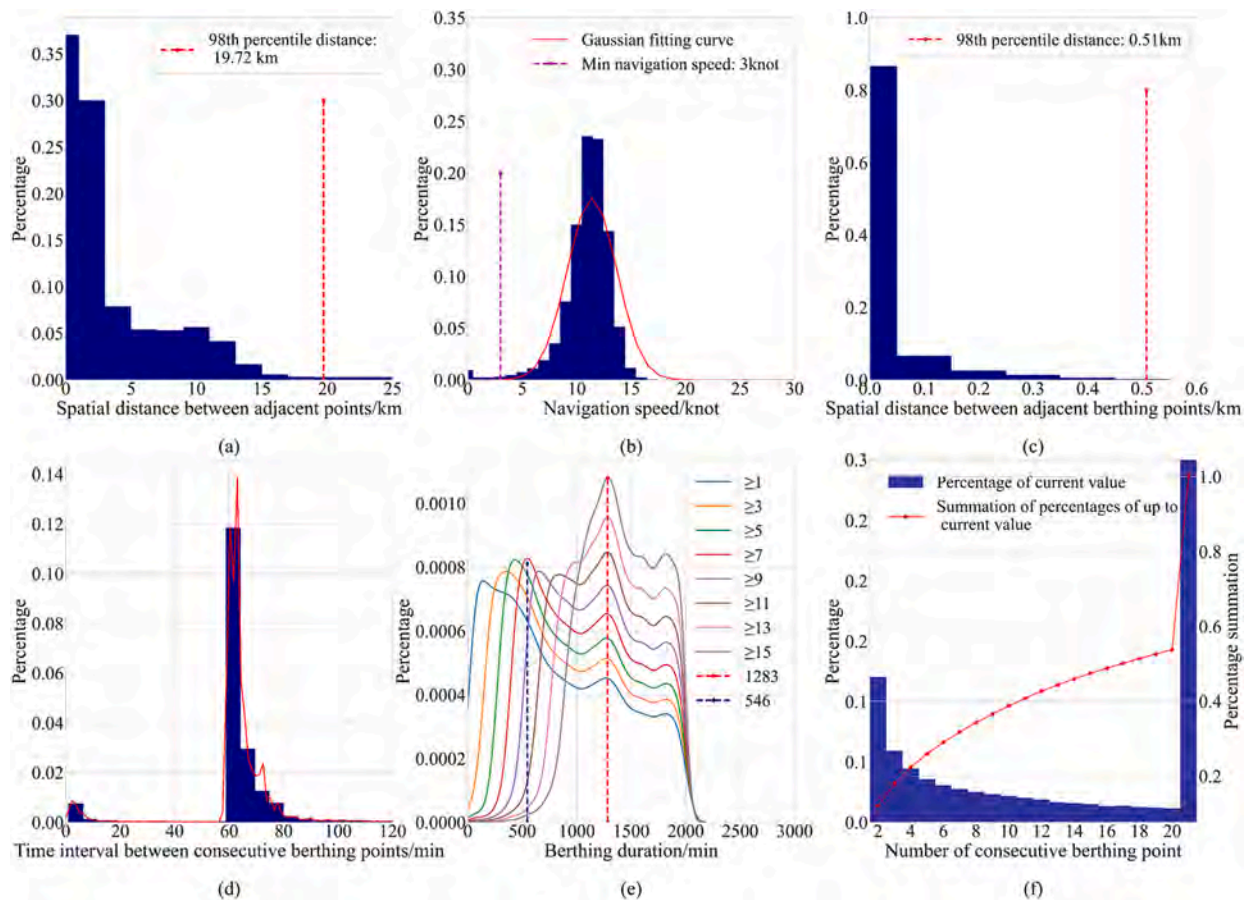


Fig. 6. Determination of vessel trajectory segmentation parameters: (a) statistical distribution of spatial intervals between all vessel points; (b) statistical distribution of vessel navigation speed; (c) statistical distribution of spatial distance between continuous berthing points; (d) statistical distribution of time intervals between continuous berthing points; (e) statistical distribution of stay duration under different continuous berthing points; (f) statistical proportion of different continuous berthing point scenarios.

Table 2 presents the content of the experimental dataset after each processing step. In the waypoint area recognition phase, not all segmented trajectories are utilised. If a vessel trajectory only operates within local waters (defined as having a range of less than 2° in both longitude and latitude directions), its waypoints are generally meaningless for route searching. Additionally, if a trajectory is lengthy but contains too few points, it indicates incomplete trajectory information, making it challenging to accurately identify waypoints. These trajectories are also excluded. Ultimately, the number of segmented trajectories is reduced from 489,395 to 108,944, but this reduction only account for a 12 % decrease in the total number of AIS points, indicating that the vast majority of trajectory data is retained.

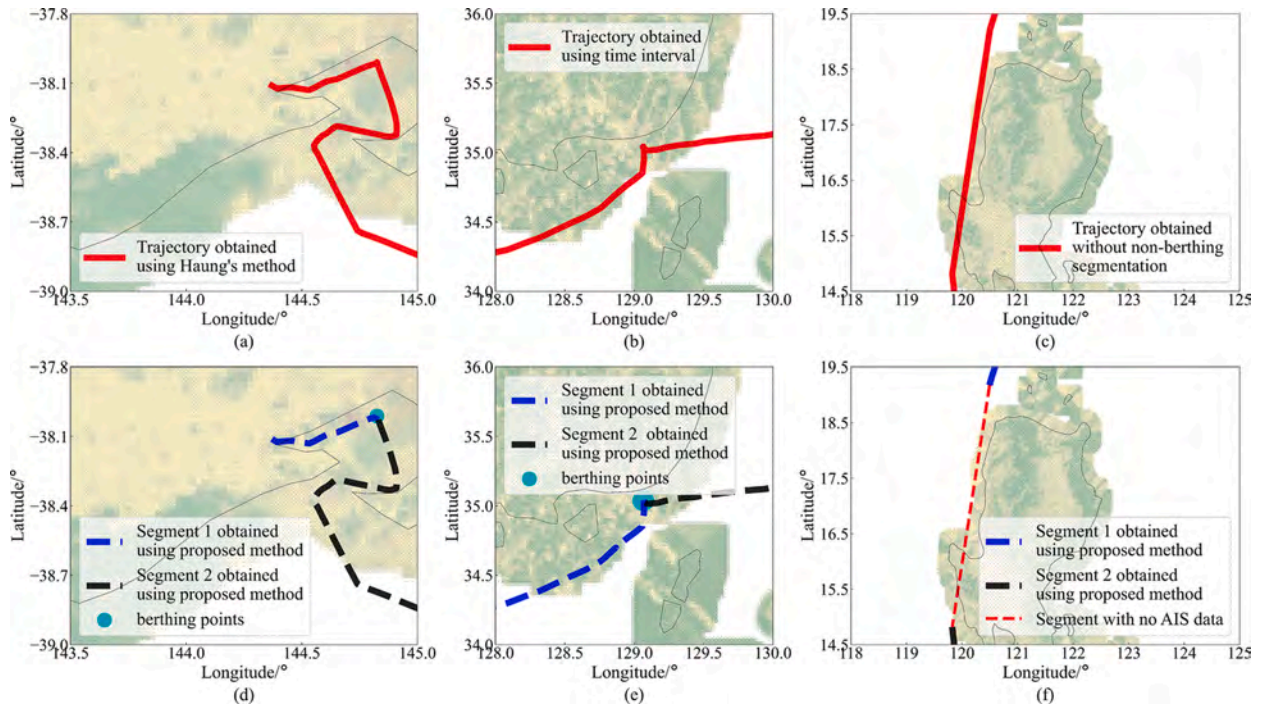


Fig. 7. Trajectory segmentation comparison of between different methods.

3.1. Vessel trajectory segmentation

3.1.1. Segmentation parameter determination

According to Algorithm 1, the first step is to determine the five parameters through statistical analysis, as illustrated in Fig. 6. Fig. 6(a) shows the statistical distribution of the spatial distances between all adjacent AIS points. The results indicate that most spatial distances fall within 15 km, with the 98th percentile distance of 19.72 km selected as the value of D_{conn} . To determine the value of T_{conn} , the speed of points under navigation status is analysed, as presented in Fig. 6(b). The sailing speed is primarily distributed between 3 and 15 knots, and a sailing speed of 3 knots is selected as S_{min} . Consequently, T_{conn} at this speed is calculated to be 213 min.

AIS Points with statuses of “At anchor” and “Moored” are used as preliminary berthing points to determine parameters related to berthing identification. These points are not directly regarded as definitive berthing points, as the status information cannot be guaranteed to be completely accurate (as shown in Fig. 8). The value of D_{stop} is determined to be 0.51 km based on the statistical distribution of spatial distances between these adjacent points, as shown in Fig. 6(c). To determine N_{stop} , Fig. 6(f) displays the distribution of the number of consecutive berthing points, which exhibits significant variations and lacks distinguishing characteristics that can differentiate temporary berthing from long-term berthing. To address this, further statistical analysis is conducted on the berthing duration under various scenarios with different berthing points, ranging from 0 to 2000 min, as shown in Fig. 6(e). The results indicate that the proportion of the first peak decreases significantly when N_{stop} exceeds 7, suggesting the possibility of vessel transitioning from temporary to long-term berthing. In combination with Fig. 6(d), which shows that most berthing points have a time interval of 60 min, seven consecutive berthing points correspond to a time interval of 420 min. Additionally, the first peak value (546 min) is identified in the curve labelled 7 in Fig. 6(e). Accordingly, N_{stop} is set to 7, while T_{stop} is set to 546 min.

3.1.2. Comparison of trajectory segmentation methods

Using these parameters, trajectory segmentation is conducted. Fig. 7 compares the segmentation results using different methods. In Fig. 7(a), AIS devices shut down at the onset of berthing without returning data, rendering Huang’s method (Huang et al., 2021) incapable of identifying these discontinuous berthing data. Although a common method involving a larger time interval threshold can mitigate this issue (Liu et al., 2023c), the relatively smaller time intervals between adjacent points in continuous berthing still necessitate consideration of the number of berthing points to segment the trajectory; otherwise, it may be considered as a single trajectory, as shown in Fig. 7(b). Finally, for AIS data collected by satellites, data loss is frequent, and without considering non-berthing segmentation, trajectories may traverse land, as shown in Fig. 7(c). In contrast, the proposed method successfully segments the trajectories in these scenarios, as demonstrated in Fig. 7(d), 7(e), and 7(f). This highlights the proposed method’s specific advantage in accurately segmenting trajectories despite data discontinuities and data loss, making it highly effective for diverse and challenging scenarios.

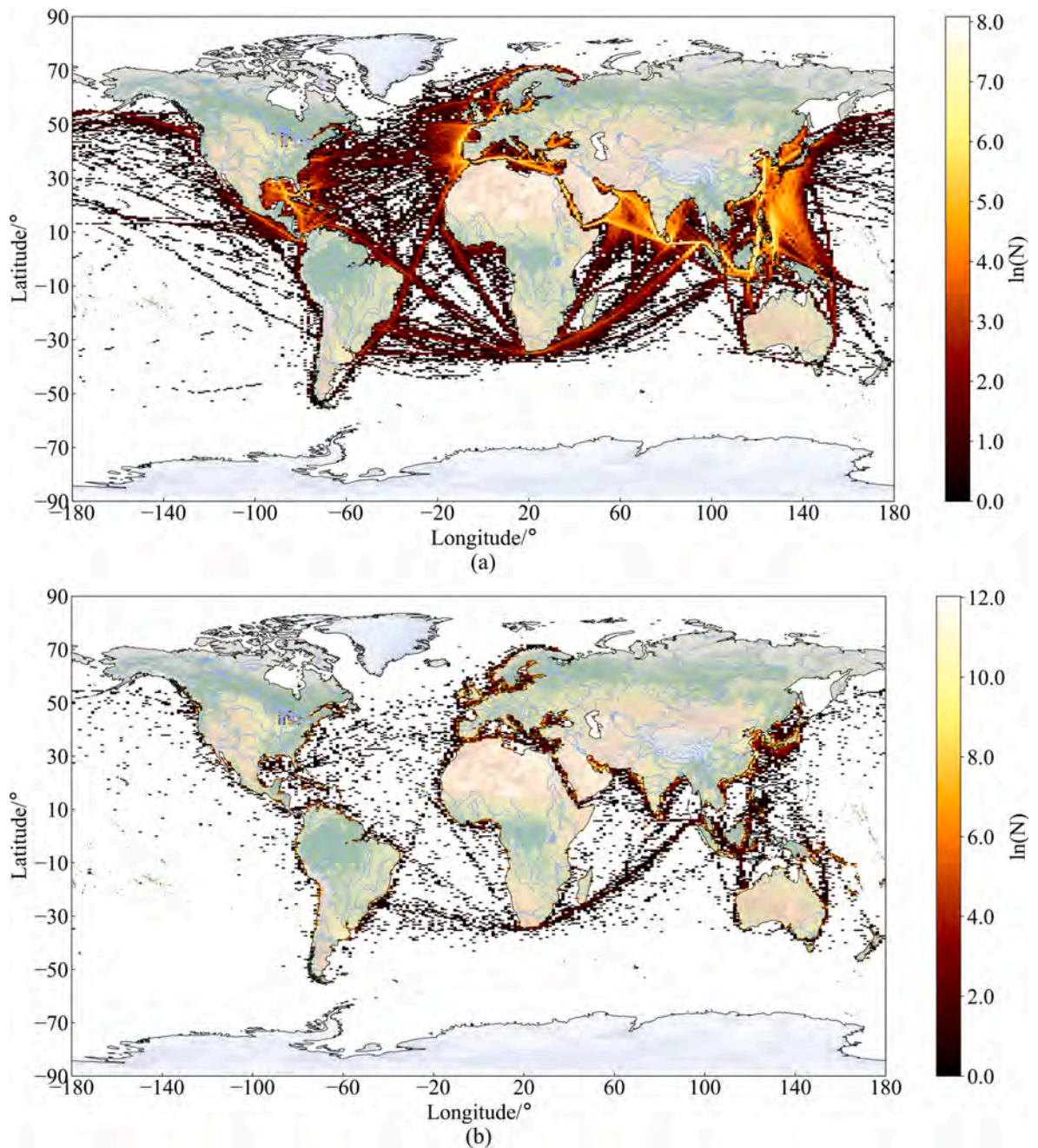


Fig. 8. Heatmaps of trajectory segmentation points. (a) trajectory-discontinuous non-berthing segmentation points; (b) berthing segmentation points.

3.1.3. Analysis of segmentation point distribution

The algorithm and thresholds described above are employed to identify berthing and non-berthing segmentation points, resulting in 489,395 segmented vessel trajectories based on berthing and non-continuous trajectory segmentation. The distribution of the 408,950 non-berthing segmentation points that are depicted in Fig. 8(a) is primarily located along major shipping routes. It is found that 98.17 % of them are in the “under way” navigation status. The distribution of the 7,094,086 berthing segmentation points, depicted in Fig. 8(b), is significantly higher than that of non-berthing segmentation points, with most concentrated near the coastline. A small number of berthing segmentation points are also distributed along shipping routes, due to prolonged anchorage periods during the voyage. Among the berthing segmentation points, 47.46 % and 45.43 % are found to be in “moored” and “at anchor” status, respectively, while 6.47 % of the data are in “under way” navigation status. This can be attributed to the fact that when a vessel is identified as being in berthing status, but the data does not return to the berthing status.

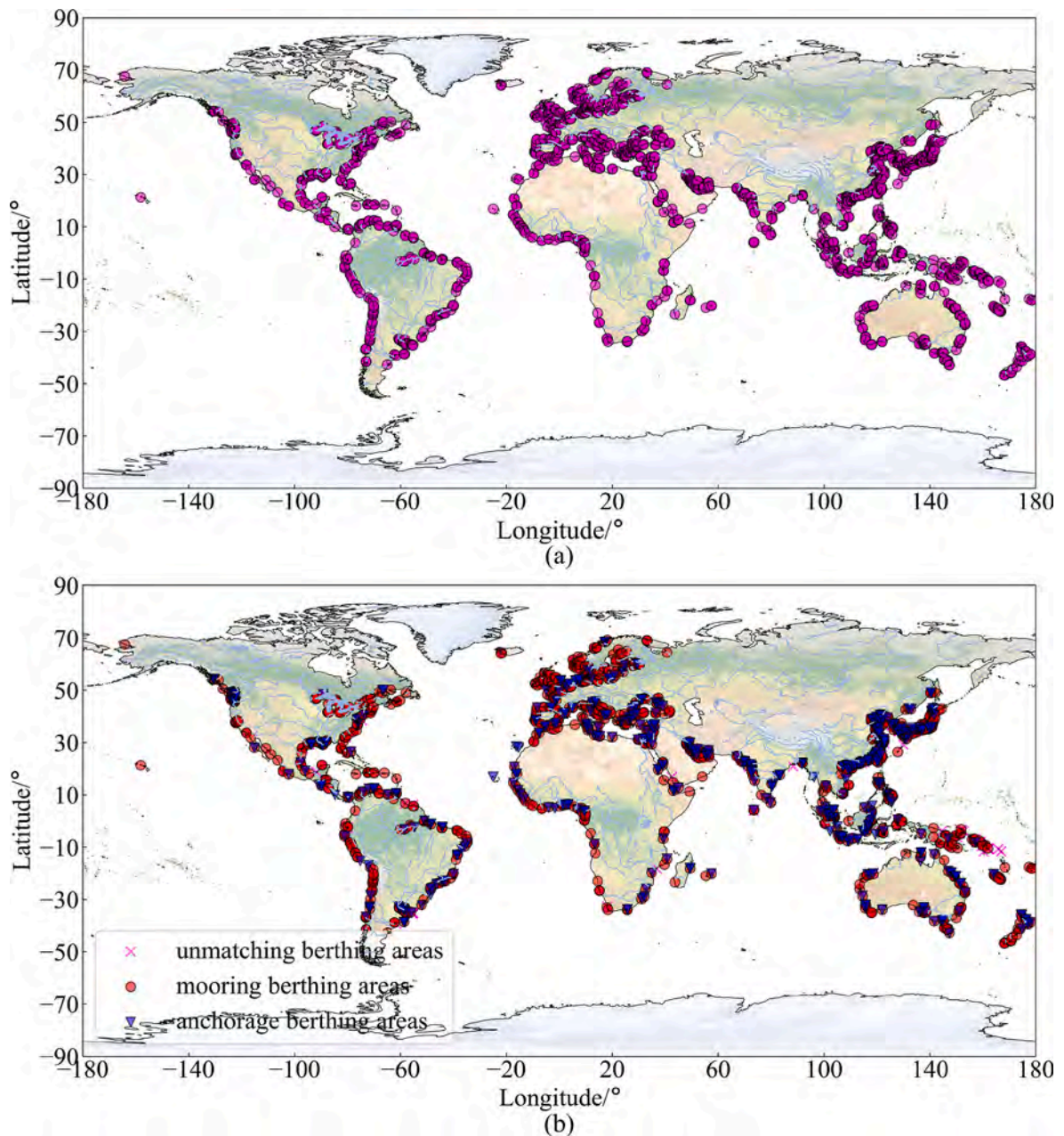


Fig. 9. Distribution of estimated berthing areas and the matching results with actual ports and anchorages. (a) distribution of estimated berthing areas; (b) the matching results with actual ports and anchorages.

3.2. The features extraction of the global maritime shipping network

3.2.1. Berthing area identification

After obtaining the berthing points as shown in Fig. 8(b), KNN-based point filtering is employed to remove 144,826 invalid berthing points with $k = 15$ and $D_k = 0.51$ km according to Fig. 6(c). The global waters are divided into 1800×900 grids with a grid size of 0.2° . Since the grid division scheme is uniformly applied for constructing berthing areas, waypoint areas, and grid networks, it should not be too fine, as this may impede effective clustering into valid blocks. Conversely, if the grid is too coarse, it may fail to adequately capture the complexities of restricted and coastal waterways. A grid size of 0.2° is chosen because it corresponds to approximately 22 km at latitude and ranges from 15 km to 22 km at varying longitudes. According to Fig. 6(a), the distances between consecutive neighbouring points typically do not exceed 15 km, making the 0.2° grid appropriate for ensuring that trajectory points remain within a single grid. This continuity is vital for forming cohesive waypoint areas and enabling smoother arcs within the grid networks. Based on this, the

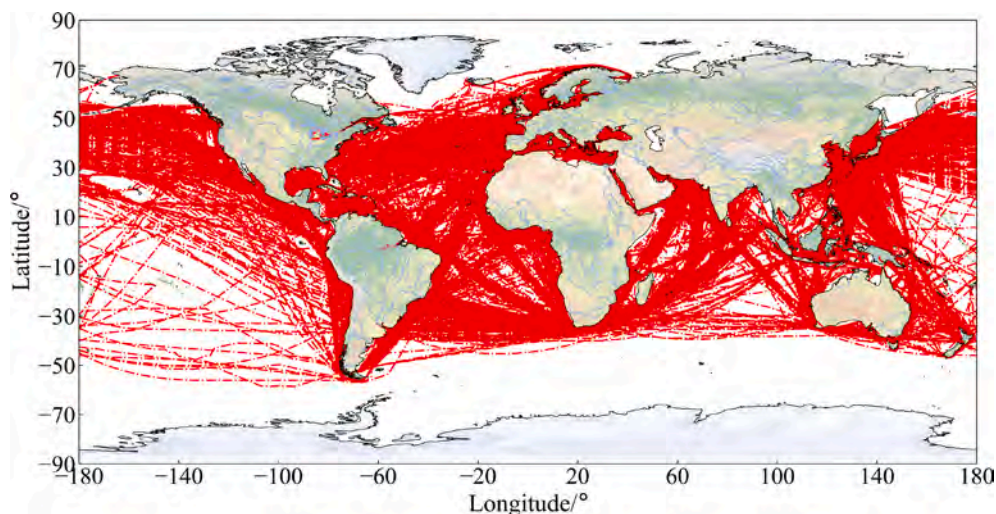


Fig. 10. Trajectory map between different estimated berthing areas.

approach mentioned in Section 2.2.1 is used to obtain 1,135 block areas.

To achieve effective clustering, the ϵ value is set within a range of 1,000 m to 5,000 m, based on two considerations. Firstly, according to Fig. 6(c), the distance between berthing points generally does not exceed 0.51 km. Taking into account the possible drifts, the lower limit was set to 1,000 m. Secondly, to prevent anchor areas and operation berthing areas from being grouped together due to an excessively large upper limit, the upper limit is restricted at 5,000 m. With $minPts = 15$, adaptive DBSCAN clustering is performed on each block area, resulting in 2,221 berthing areas, as shown in Fig. 9(a). The berthing points forming these berthing areas account for 96.79 % of all berthing points.

To further validate the identification results of berthing areas, global ports and anchorages data are extracted from the Seasearcher website,¹ including information on area categories, names, and location. Since the precise latitude and longitude of the given areas cannot be obtained, an assumed distance threshold of $d_{port} = 20km$ is applied. A berthing area is considered a match with an actual location if the distance from its centre to a port or anchorage did not exceed d_{port} . As a result, 1,470 berthing areas are matched with ports, 610 are matched with anchorages, and 141 cannot be matched, as illustrated in Fig. 9(b). Most of the berthing areas are successfully matched with actual ports or anchorages, demonstrating the effectiveness of the berthing point identification method and the adaptive clustering method to some extent.

3.2.2. Construction of navigable routes database

Subsequently, vessel trajectories are matched with identified berthing areas, resulting in obtain 293,721 historical trajectories between berthing areas. From this dataset, the shortest trajectory for each pair of berthing areas is selected, forming a navigable route database comprising 20,943 routes, as shown in Fig. 10. These trajectories serve as a foundation for data-driven vessel route matching.

3.3. Waypoint identification and shipping network generation

3.3.1. Comparison of waypoint identification approaches

To validate the reliability of the waypoint detection algorithm in waypoint identification, Fig. 11 presents a comparison of the results obtained using threshold detection (Liu et al., 2023c), local outlier factor anomaly detection, DP trajectory compression detection (Wei et al., 2020; Zhang et al., 2018), and waypoint detection on the same trajectory. It is evident that threshold detection yields excessive waypoints and may misidentify intermediate turning points as waypoints. Although local outlier factor detection and DP trajectory compression detection reduce the number of detected waypoints, it may also mistakenly identify intermediate turning points as waypoints and fail to identify some actual turning points. In contrast, the waypoint detection algorithm performs better in identifying waypoints. It avoids designating all points as waypoints when the heading is unstable and effectively controls the number of detected waypoints. According to Equation (4), the results indicate that threshold detection achieved a score of 308, local outlier factor detection scored 161, DP trajectory compression attained a score of 136, and the proposed method reached a score of 80, further demonstrating its superiority. Therefore, the PELT algorithm is employed to identify waypoints for all segmented trajectories. As a result, a total of 430,007 waypoints are obtained.

¹ Lloyd's List Intelligence: <https://www.lloydslistintelligence.com/>.

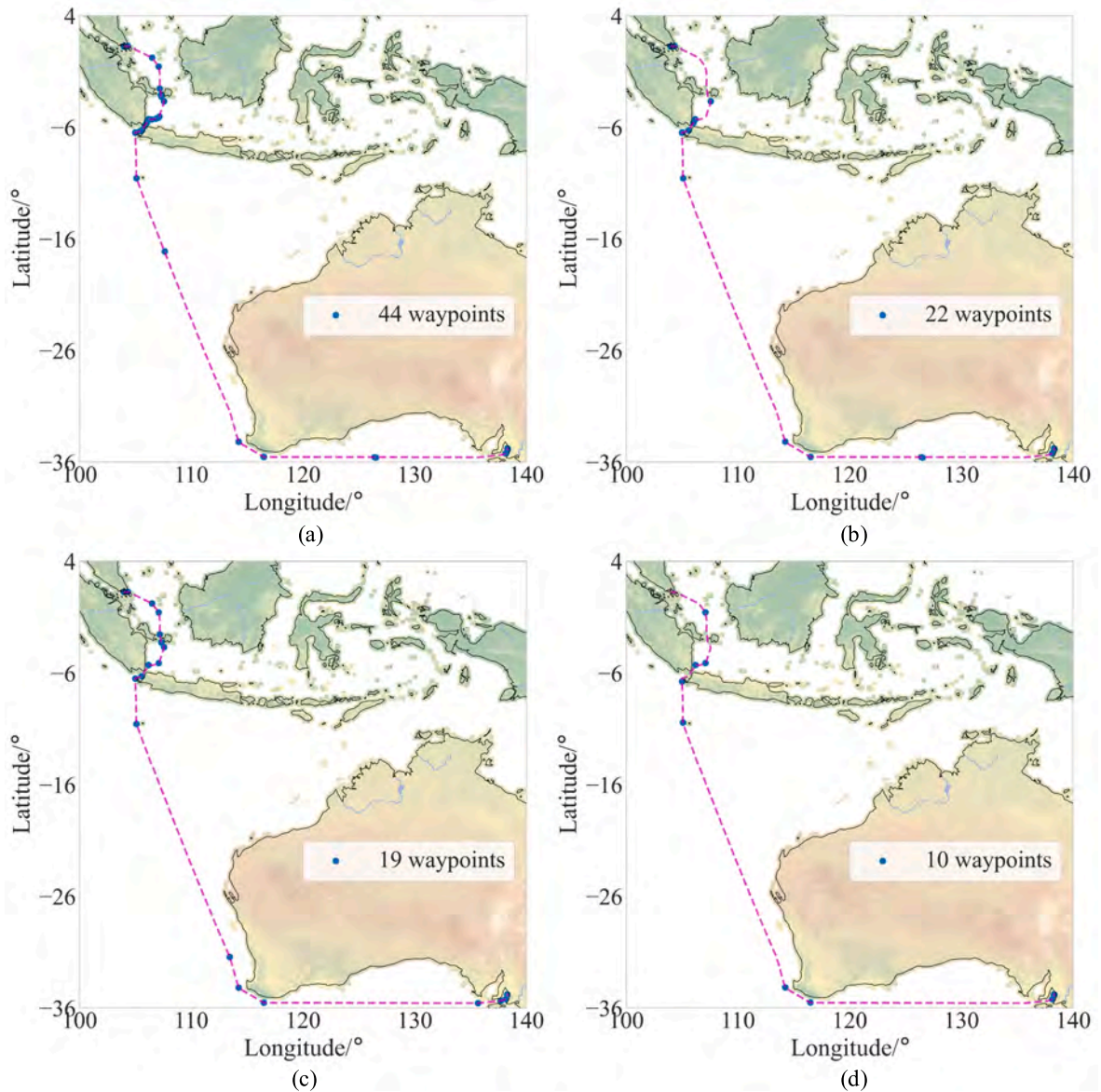


Fig. 11. Comparison of different methods for detecting waypoints. (a) detection result based on a heading difference threshold of 10 degrees; (b) detection result of local outlier factor anomaly detection; (c) detection result of the DP compression algorithm with a distance threshold of 10 km; (d) detection results of the waypoint detection by the PELT algorithm.

3.3.2. Waypoint area identification

The CKBA-DBSCAN clustering method is still used for waypoint area identification. Fig. 12 illustrates the waypoint distribution before and after KNN filtering. Due to the absence of fixed shipping channels and the unique behaviour of vessels, certain regions where only a small number of vessels perform manoeuvres fail to form conventional waypoint areas, as shown in Fig. 12(a). Retaining these waypoints also result in the majority of grid cells being considered active and connected together after the water area is partitioned, making it difficult to effectively form blocks. Consequently, the KNN algorithm is employed for waypoint filtering, with the value of k set to 15, representing the minimum number of waypoints required to form a waypoint area. Considering the characteristics of effective waypoint aggregation, a curve showing the relationship between the distance threshold and the proportion of remaining waypoints is plotted. Based on the growth rate variation, the distance threshold is set to 30,000 m. After filtering, 373,243 waypoints remain, as showed in Fig. 12(b).

The global waters are divided into 1800×900 grids, and 1,145 valid block areas are obtained. For the adaptive DBSCAN clustering, with $minPts$ set to 15, the ϵ value range is defined as [0.01, 0.2] after normalising both directional and spatial distance using min-max normalisation and computing their weighted sum. Fig. 13 presents the results of adaptive clustering and fixed-parameter clustering of waypoints within different blocks. In Fig. 13(a), the optimal ϵ value for Block 362 is 0.14. With this parameter, two effective clusters

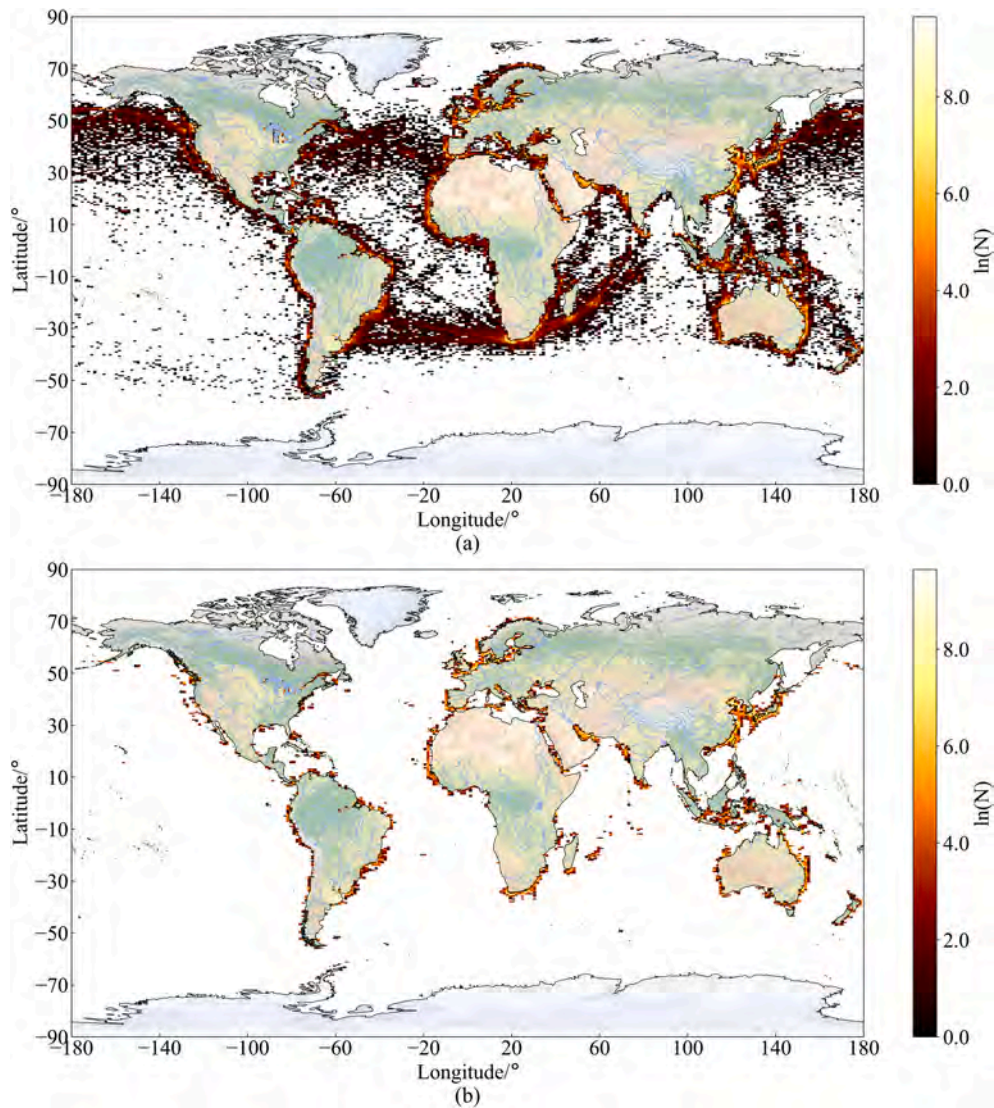


Fig. 12. Waypoint distribution. (a) raw waypoint distribution; (b) KNN-filtered waypoint distribution.

are obtained, while a significant number of waypoints in Block 224 are classified into the three cluster under the same parameter setting. Conversely, the optimal epsilon for Block 224 is 0.01. With this parameter, a total of 66 effective clusters are obtained, but none of the waypoints in Block 362 can form a waypoint area using this parameter. This indicates the difficulty in unifying clustering parameters for such large-scale water bodies with significant differences in waypoint density among regions. And the proposed method effectively addresses this challenge by employing block-specific parameters, ensuring accurate and efficient clustering across diverse maritime regions. Additionally, the vessel course characteristics of waypoints mean that spatially close points may still be assigned to different clusters. For instance, in Fig. 13(c), clusters 1 and 3 are separated from cluster 2, and in Fig. 13(f), clusters that are close or even overlapping are distinguished.

After performing block adaptive clustering on all waypoints, 2,588 waypoint areas are identified. These waypoints accounted for 81.6 % of the recognised waypoints. However, some trajectories have too few points to form a waypoint region in the identified region, while in other cases, a vessel's manoeuvring behaviour in a local area causes it to deviate from other vessels on the same route, making it difficult to form a waypoint region. Fig. 14(a) shows that the waypoint areas are mainly distributed near the coastline and on the main shipping routes. Although the change point detection method avoids the problem of recognising too many trajectory points close to ports as waypoints using a threshold detection method, multiple vessels still exhibit turning behaviours when approaching these areas.

3.3.3. Construction and analysis of the maritime shipping network

The nodes of the shipping network are derived by merging the berthing areas and waypoint areas, while the trajectory segments are

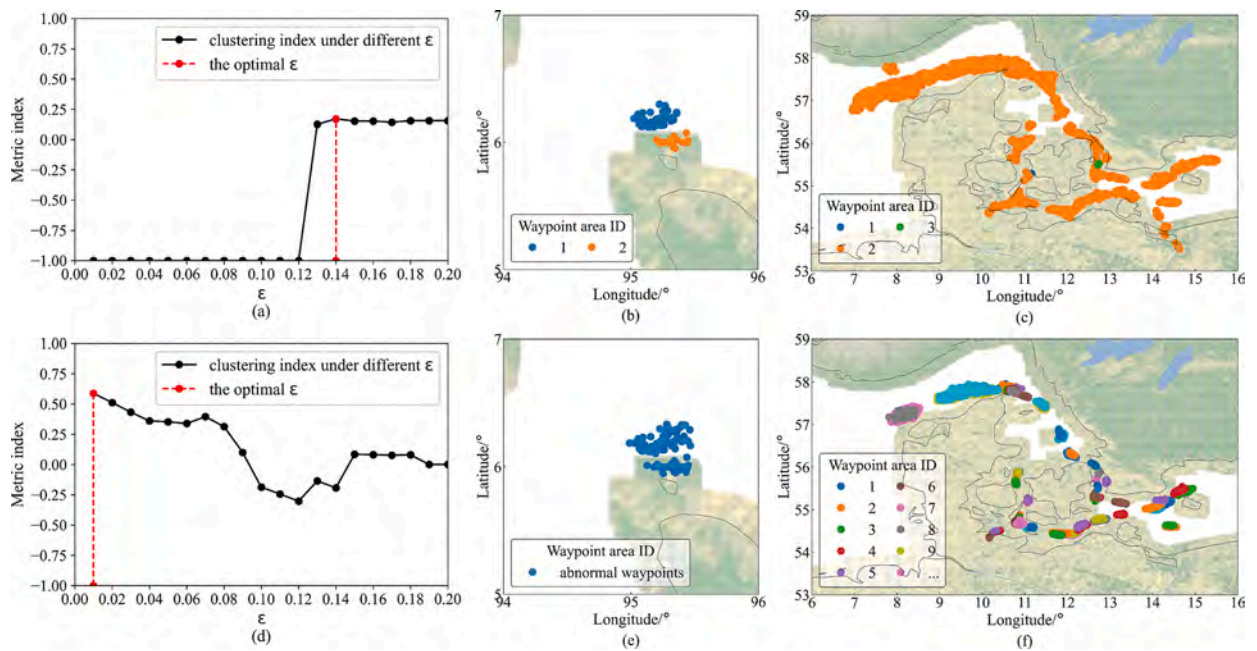


Fig. 13. Comparison of clustering results under different ϵ . (a) clustering index curve for Block 362; (b) adaptive clustering results for Block 362; (c) clustering results for Block 224 under optimal parameters for Block 362; (d) clustering index curve for Block 224; (e) clustering results for Block 362 under optimal parameters for Block 224; (f) adaptive clustering results for Block 224.

utilised to generate connecting arcs between nodes. Finally, the shipping network is constructed, as shown in Fig. 14(b). For navigation routes with very few trajectories, where the number of waypoints on the trajectories is insufficient to form waypoint areas, these waypoints are treated as general navigation points. It can also be seen that the trajectories between nodes have a certain degree of overlap, leading to nodes on the same route being covered by other trajectories. This is because trajectories on different routes may exhibit different patterns when passing through the same area.

We further analyse the shipping networks, as shown in Fig. 15, where the values indicated by the colour bar represent the degree of nodes or the number of arcs between nodes. Fig. 15(a) displays the distribution of nodes with different degrees. Nodes with higher degrees are mainly distributed near the coastline, and dry bulk cargo vessels prefer the Cape of Good Hope route instead of the Suez Canal route combined with the further analysis of Fig. 15(b). It also can be found from Fig. 15(b) that dry bulk cargo shipping primarily operated around East Asia, Australia, and Brazil.

3.3.4. Development of the grid network

To generate the grid network, the global water areas are also divided into 1800×900 grid system. Trajectory segments are then serialised into grid sequences, with grid connectivity determined by each pair of adjacent points within the segments. Furthermore, the directionality of the segments is preserved, ensuring that the grid network remains a directed graph.

3.4. Comparison studies

3.4.1. Comparative analysis of navigable route generation

This study compares navigable routes generated by three different methods: the trajectory searching method, the maritime shipping network method, and the double-network matching method. Initially, the analysis focuses on scenarios where navigable routes are identified through trajectory matching. Fig. 16 compares navigable route planning results in two distinct trajectory scenarios. The authors observe that in Fig. 16(a), the path generated based on the double-layer network matching mainly aligns with the path generated by the shipping network, whereas the trajectory searching result is longer than the former two. In Fig. 16(b), the routes generated based on the shipping network and the double-layer network matching are substantially similar, but they significantly differ from the trajectory searching result.

Table 3 provides a detailed route planning results across two scenarios. From the perspective of time complexity, the trajectory matching method yields results quickly as it only requires data matching. In contrast, double-layer networks searching approach involves a search process at the grid network level before proceeding to matching the maritime shipping network, leading to significantly longer computation times. However, in terms of route length, the network-based search demonstrates superior performance, underscoring the potential of the proposed method for application in route planning. Nonetheless, this advantage may partly arise from the limitation of not accounting for the influence of weather conditions.

Furthermore, when navigable routes cannot be found through trajectory searching, the shipping network and double-layer network

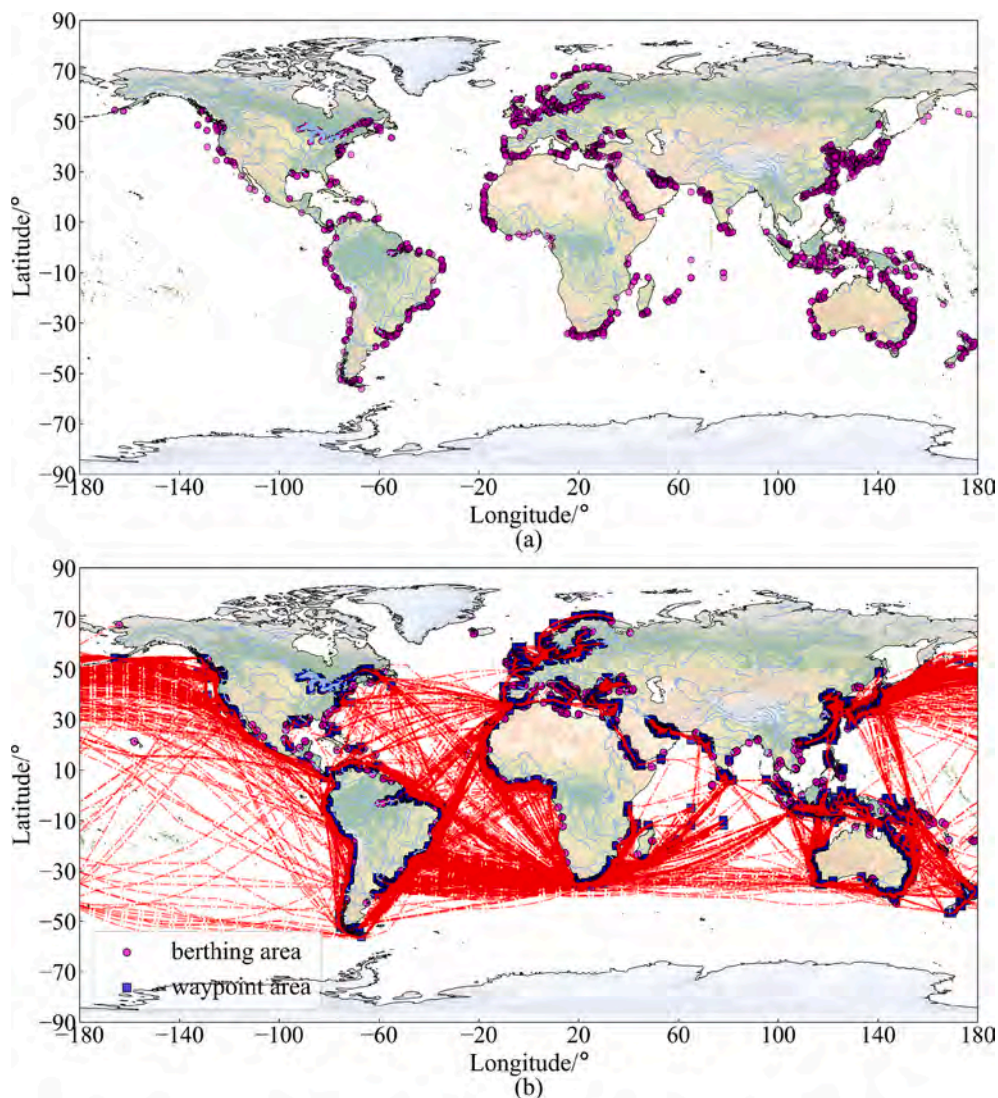


Fig. 14. Results of waypoint area recognition and estimated maritime shipping network. (a) results of waypoint area recognition; (b) estimated maritime shipping network.

matching can be used to generate navigable routes. As shown in Fig. 17(a), the berthing area in the maritime shipping network are considered searchable intermediate nodes, resulting in local anomalies in the generated path. Conversely, the path generated using the proposed double-layer network matching is found to be superior. Additionally, due to limited historical trajectories, the searchable path range of the shipping network is constrained, as illustrated in Fig. 17(b). Nevertheless, the double-layer network matching can still successfully obtain a navigable route between the starting and ending points.

To verify the effectiveness of the proposed method for other types of vessels, the container vessel AIS data from 2021 is used to construct the maritime shipping networks and perform route planning, as shown in Fig. 18. The diversity of the original trajectories leads to a certain degree of instability in trajectory searching methods, resulting in navigable paths that may not be the shortest. Similarly, methods based on maritime shipping networks can yield abnormal results due to the limited accessibility of nodes generated from trajectories and the treatment of berthing areas as intermediate nodes. In contrast, the paths obtained through double-layer networks are more optimal. The network construction process and path search method using container vessel AIS data further validate the effectiveness of the proposed framework and methods.

3.4.2. Comparative analysis of route planning performance

To demonstrate the effectiveness of the proposed method, Table 4 further compares the route planning results from the same departure port to different destination ports using various methods. These methods include existing route planning software and current trajectory searching techniques. Specifically, the existing voyage optimisation tool is employed to generate port-to-port connections. This tool automatically optimises routes by taking into account vessel navigation requirements and channel

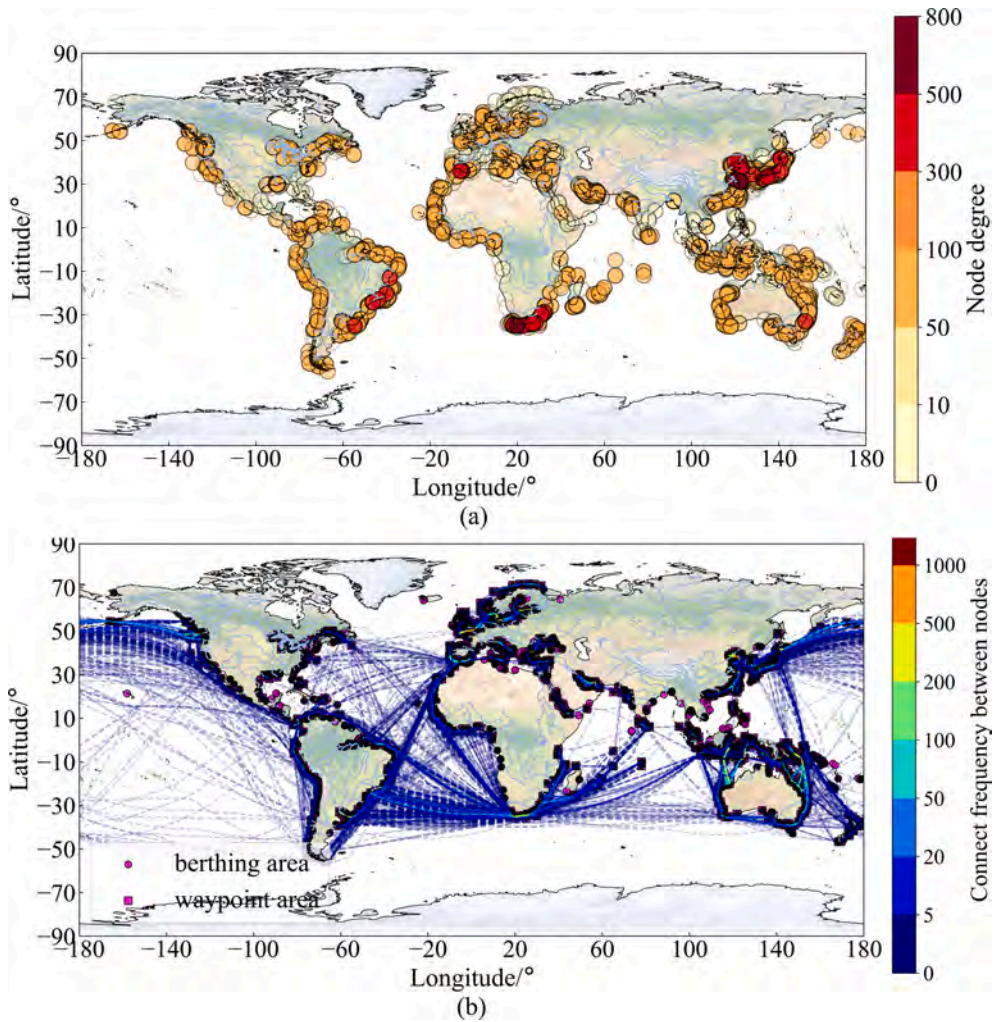


Fig. 15. Analysis of maritime shipping networks. (a) network node degree analysis; (b) network link frequency analysis.

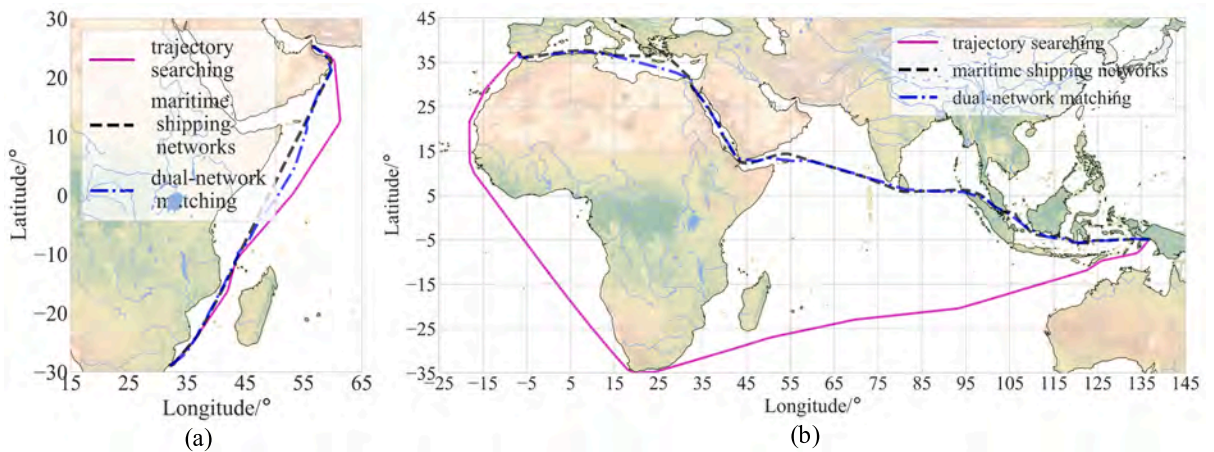


Fig. 16. Comparison between original trajectories and network-based generated routes. (a) short-distance shipping scenario; (b) long-distance shipping scenario.

Table 3

Comparison of route planning results across two scenarios in Fig. 16 using different methods.

	Route planning methods	Searching time (s)	Route length (km)
(a)	trajectory matching	<0.001	7255.6
	maritime shipping network searching	1.625	6844.3
	double-layer networks searching	15.945	6963.6
(b)	trajectory matching	<0.001	22231.1
	maritime shipping network searching	59.498	17045.7
	double-layer networks searching	283.053	16854.3

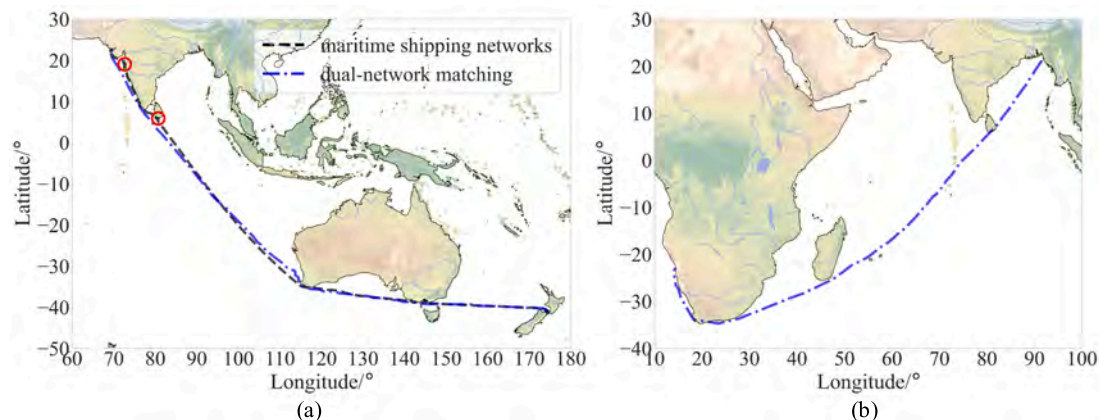


Fig. 17. Route generation in the absence of historical sailing trajectories. (a) scenarios for generating navigable routes by both maritime shipping networks and double-layer networks; (b) scenarios for generating navigable routes only using double-layer networks.

restrictions, based on the input of departure and destination port information. Fig. 19 includes five globally representative shipping routes. The results of global route planning indicate that the method proposed in this study can efficiently and rapidly generate and plan these routes.

- **Route 1: Shanghai Port to Shantou Port.** This route connects the bustling metropolis of Shanghai, one of China's major economic centres and seaports, with Shantou Port, located in Guangdong Province. Shantou Port serves as an important gateway for trade in the South China Sea region, facilitating the transportation of goods and materials to and from southern China.
- **Route 2: Shanghai Port to Port of Singapore.** This route links Shanghai Port, a key hub for international trade and commerce in East Asia, with the Port of Singapore, one of the world's busiest and most strategically located ports. The Port of Singapore serves as a crucial transshipment hub, connecting major shipping routes between Asia, Europe, and the rest of the world.
- **Route 3: Shanghai Port to Port of Cape Preston.** This route connects Shanghai Port with the Port of Cape Preston, located in Western Australia. The Port of Cape Preston primarily serves the iron ore mining industry in the Pilbara region, facilitating the export of iron ore to global markets, particularly to Asia.
- **Route 4: Shanghai Port to Port of Itaguaí.** This route connects Shanghai Port with the Port of Itaguaí, situated in Rio de Janeiro, Brazil. The Port of Itaguaí is a major port for the export of iron ore and other commodities, serving as a crucial link in Brazil's maritime trade network and facilitating trade with international markets.
- **Route 5: Shanghai Port to Port of Rotterdam.** This route connects Shanghai Port with the Port of Rotterdam, Europe's largest and busiest seaport located in the Netherlands. The Port of Rotterdam serves as a vital gateway for European trade, handling a diverse range of cargo including containers, bulk goods, and petroleum products, and facilitating trade between Europe and the rest of the world.

Route planning results based on maritime shipping networks are excluded from this comparison due to potential anomalies. Table 4 demonstrates the superior route searching speed of the proposed method, especially when historical trajectories are available, resulting in significantly faster route searches. Table 4 also highlights that those longer routes do not necessarily result in slower search speeds. For example, the journey from Shanghai Port to the Port of Itaguaí is longer than to the Port of Rotterdam, yet the search time is shorter. This discrepancy is attributed to the complexity of maritime environments traversed, where routes passing through open waters require less time for searching compared to those navigating through complex maritime passages such as the Malacca Strait and the Suez Canal.

More specifically, the proposed method, utilising double-layer networks searching, consistently achieves significantly lower searching times compared to software searching. For example, from Shanghai Port to Shantou Port, the searching time is reduced from over 480 s to just 0.75 s, showcasing a remarkable improvement in efficiency. Additionally, the proposed method consistently

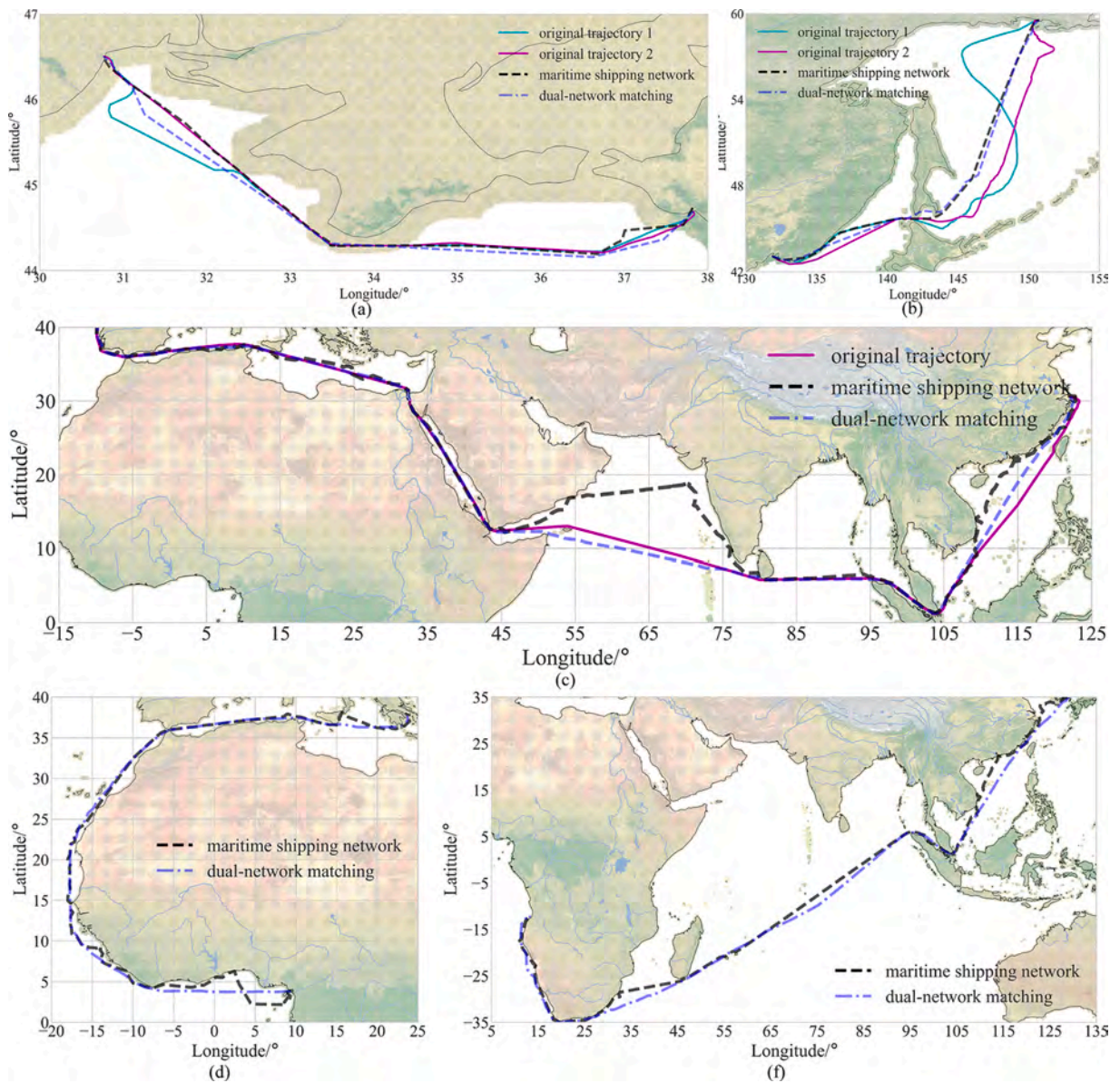


Fig. 18. Route searching and planning generation based on the container vessel AIS data. (a) local navigable path search with original trajectories; (b) short-distance navigable path search with original trajectories; (c) long-distance navigable path search with original trajectories; (d) short-distance navigable path generation without original trajectories; (e) long-distance navigable path generation without original trajectories; (f) long-distance navigable path generation without original trajectories.

produces optimized route lengths compared to software searching. For instance, from Shanghai Port to Shantou Port, the route length is reduced from 1279.1 km to 1137.9 km, indicating more efficient route planning with the proposed method. Moreover, the proposed method demonstrates consistent performance across different destination ports, consistently outperforming software searching in terms of searching time and route length optimisation. This suggests the robustness and reliability of the proposed method in various scenarios, enhancing its applicability in diverse maritime contexts.

3.5. Discussions and implications

This study presents a novel approach to grid generation for the global maritime network, designed to facilitate automatic vessel route planning by leveraging AIS data. The primary contribution of this research lies in the innovative use of historical big data to construct a structured grid representation of the global maritime network. Building upon this foundation, the proposed route planning methodology enhances the accessibility and usability of the network, enabling more efficient and reliable route planning for large-scale maritime operations.

Table 4

Comparison of route planning results from the same departure port to different destination ports under different methods.

Departure port	Destination port	Route planning methods	Searching time (s)	Route length (km)
Shanghai Port	Shantou Port	trajectory matching	/	/
		double-layer networks searching	0.75	1137.9
		software searching	>480	1279.1
	Port of Singapore	trajectory matching	<0.001	4062.3
		double-layer networks searching	4.1	3961.0
		software searching	>480	4130.4
	Port of Cape Preston	trajectory matching	<0.001	5305.8
		double-layer networks searching	13.2	5905.6
		software searching	>480	6080.3
	Port of Itaguaí	trajectory matching	/	/
		double-layer networks searching	152.3	20299.2
		software searching	>480	21006.9
	Port of Rotterdam	trajectory matching	/	/
		double-layer networks searching	398.1	19137.6
		software searching	>480	19778.1

From a theoretical perspective, this study offers significant contributions to the efficient and accurate processing of large-scale AIS data. First, the proposed spatiotemporal segmentation approach proves highly effective in addressing diverse vessel trajectory segmentation scenarios. By reducing reliance on predefined segmentation parameters, this method achieves greater flexibility and adaptability, paving the way for broader applications in vessel itinerary analysis. Second, the integration of change-point detection to account for vessel heading variations marks an advancement over traditional threshold- and compression-based methods, providing a robust framework for identifying vessel behavior. By converting feature data into time series representations, this method expands its applicability to the motion analysis of both vessels and other vehicular systems. Third, complementing existing parallel processing methods for large-scale AIS data management (Li et al., 2024), the study introduces a block-based adaptive processing strategy that effectively addresses the challenges posed by extensive and heterogeneous datasets. This approach demonstrates exceptional adaptability across varying maritime regions, making it particularly suitable for global-scale data-driven maritime research and practical applications.

From a practical standpoint, this study also delivers a robust route planning methodology tailored for long-distance oceanic navigation. Comparative experimental analyses between the proposed method and the existing voyage optimisation tool underscore the method's advantages in significantly reducing computational time and optimizing route lengths. These benefits arise from the grid-based global maritime network constructed from historical vessel trajectories, which narrows the search area and enables rapid route identification. Unlike traditional methods that involve exhaustive exploration of all potential areas between ports, the proposed approach capitalizes on the predefined network structure to achieve faster and more efficient route planning.

Despite its demonstrated advantages in AIS data processing and route searching, opportunities remain for further refinement to enhance its practical applicability. Vessel maneuvering is inherently a continuous process, yet change-point detection typically isolates the most pronounced variation points within this process. Future research could focus on integrating change-point detection with complementary methods to capture the entire continuous maneuvering process comprehensively. Additionally, as observed in Fig. 16 (b) and 18(c), certain inconsistencies in search results arise from the inherent limitations of the maritime network's connectivity, which is based on historical vessel trajectories. These limitations constrain the network's accessibility, particularly in areas where connections between nodes are sparse or incomplete. Enhancing inter-node connectivity by leveraging the similarity of incoming and outgoing directional vectors, along with spatial relationships between nodes, could substantially improve the accessibility and robustness of the maritime network. Such improvements would, in turn, enhance the reliability and precision of the route planning results derived from the network.

4. Conclusions and future research

The study introduces an innovative method for integrating data-driven and model-based approaches to automatically vessel routes at a global level, considering all bulk carriers as main sample vessels. The integration method introduced accounts for: (1) a data mining model for trajectory segmentation, waypoint detection, and clustering of berthing points; (2) the development of a global maritime shipping network for vessel route planning from historical data; (3) the A* algorithm and a double-layer network matching method for generating the final navigable route for any global voyage. The method is demonstrated using data covering one year and considering all the worldwide operating bulk carriers. Key conclusions can be summarized as follows:

- The innovative integration of data-driven and model-based approaches shows promise for global route planning. By combining the strengths of both methods, it offers a comprehensive solution (see Figs. 16 and 17).
- The vessel trajectory segmentation based on the spatiotemporal distance of vessel points takes into account the differences in spatiotemporal characteristics of trajectory points under different scenarios (Algorithm 1) and can better meet the needs of vessel trajectory segmentation in complex scenarios (berthing segmentation and non-berthing segmentation).

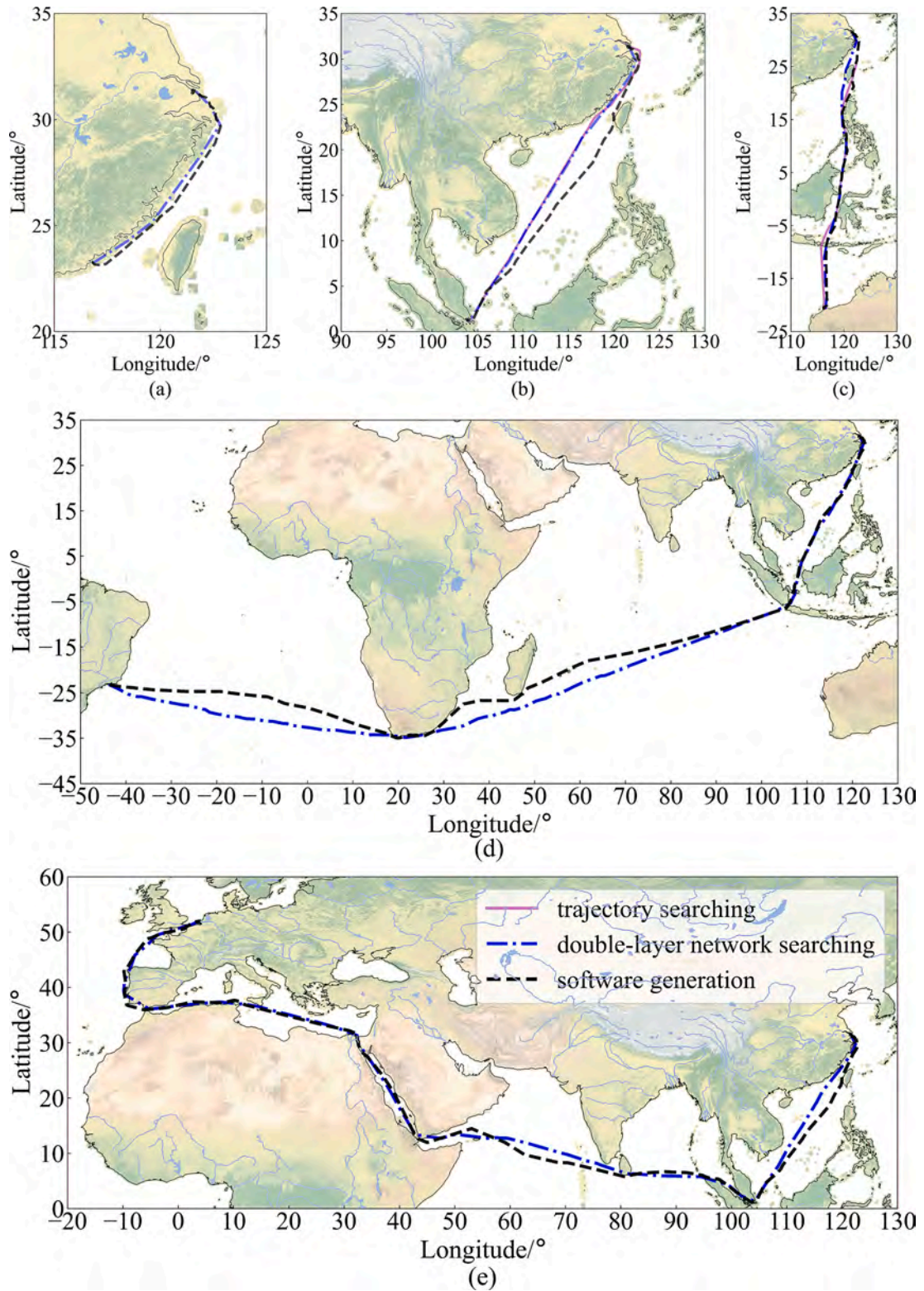


Fig. 19. Comparison of global route planning using existing route planning software and current trajectory searching techniques. (a) Route 1: Shanghai Port to Shantou Port; (b) Route 2: Shanghai Port to Port of Singapore; (c) Route 3: Shanghai Port to Port of Cape Preston; (d) Route 4: Shanghai Port to Port of Itaguaí; (e) Route 5: Shanghai Port to Port of Rotterdam.

- The proposed waypoint detection method based on the PELT algorithm is more effective at identifying global navigation points compared to the transitional threshold detection and anomaly detection methods.
- The CKBA-DBSCAN effectively constructs regions of stationary vessels and navigation points in large or even global water areas.
- Utilising global AIS data is effective for reconstructing detailed historical vessel routes, which enhances the development of data-driven route planning solutions.
- A new approach for navigable route search, utilising a double-layer network matching method, is proposed to enhance the efficiency of vessel route planning. This method broadens the accessibility between ports beyond just relying on the maritime shipping network. It also facilitates smoother route creation through grid networks, thereby increasing the probability of identifying the shortest route (Table 3).
- The proposed route planning method not only rapidly provides a global route, but also supplies suggested vessel kinematic information (such as speed, course, heading, etc.) derived from the data-driven model.

Overall, the study concludes that the search for vessel navigable routes progressively expands from matching specific trajectories to exploring navigable routes based on the shipping network and a double-layer network approach. The developed method demonstrates versatility, catering to various requirements in navigable route planning, and offers substantial support for practical route planning and shipping network analysis. For future research, it is essential to acknowledge that the planned route determined by using the proposed method is a static route with vessel kinematic information. The methods employed in this study lack consideration for various factors that influence route planning, such as vessel characteristics and hydro-meteorological conditions. There is a need to develop customized shipping networks. This entails conducting subsequent investigations focusing on route optimisation and weather routing, taking into account specific vessel characteristics and prevailing weather conditions. By addressing these factors, the route planning process can be further refined to enhance safety and efficiency. Moreover, future efforts will explore the utilisation of historical trajectory data to identify secure navigation channels, followed by optimising fuel consumption and reducing pollution emissions by integrating vessel energy consumption models into the route planning framework.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Generative AI (e.g., ChatGPT) to enhance the clarity of formulations and the quality of written language. After employing this tool, the authors carefully reviewed and edited the content as necessary, taking full responsibility for the final content of the publication.

CRedit authorship contribution statement

Lei Liu: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mingyang Zhang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Cong Liu:** Writing – review & editing, Visualization, Validation, Methodology, Investigation. **Ran Yan:** Writing – review & editing, Investigation. **Xiao Lang:** Writing – review & editing, Visualization, Validation. **Helong Wang:** Writing – review & editing, Visualization, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Mingyang Zhang acknowledges research funding received from the Horizons Europe project “RETROFIT solutions to achieve 55 % GHG reduction by 2030 (RETROFIT55) – Project No.: 101096068”. The views set out in this study are those of the authors and do not necessarily reflect the views of their sponsors.

References

- Alderson, D.L., Funk, D., Gera, R., 2019. Analysis of the global maritime transportation system as a layered network. *J. Transp. Secur.* 13 (3–4), 291–325. <https://doi.org/10.1007/s12198-019-00204-z>.
- Andersson, P., Ivehammar, P., 2017. Dynamic route planning in the Baltic Sea Region – A cost-benefit analysis based on AIS data. *Maritime Economics Logistics* 19 (4), 631–649. <https://doi.org/10.1057/mel.2016.18>.
- Bläser, N., Magnussen, B.B., Fuentes, G., Lu, H., Reinhardt, L., 2024. MATNEC: AIS data-driven environment-adaptive maritime traffic network construction for realistic route generation. *Transp. Res. Part C Emerging Technol.* 169. <https://doi.org/10.1016/j.trc.2024.104853>.
- Bellman, R., 1952. On the theory of dynamic programming. *Proc. Natl. Acad. Sci.* 38 (8), 716–719. <https://doi.org/10.1073/pnas.38.8.716>.
- Bentin, M., Zastrau, D., Schlaak, M., Freye, D., Elsner, R., Kotzur, S., 2016. A new routing optimization tool-influence of wind and waves on fuel consumption of ships with and without wind assisted ship propulsion systems. *Transp. Res. Procedia* 14, 153–162. <https://doi.org/10.1016/j.trpro.2016.05.051>.
- Cai, J., Chen, G., Lützen, M., Rytter, N.G.M., 2021. A practical AIS-based route library for voyage planning at the pre-fixure stage. *Ocean Eng.* 236. <https://doi.org/10.1016/j.oceaneng.2021.109478>.

- Chen, C., Chen, X., Ma, F., Zeng, X., Wang, J., 2019. A knowledge-free path planning approach for smart ships based on reinforcement learning. *Ocean Eng.* 189, 106299. <https://doi.org/10.1016/j.oceaneng.2019.106299>.
- Charalambopoulos, N., Xidiás, E., Nearchou, A., 2023. Efficient ship weather routing using probabilistic roadmaps. *Ocean Eng.* 273. <https://doi.org/10.1016/j.oceaneng.2023.114031>.
- Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. *Numer. Math.* 1 (1), 269–271. <https://doi.org/10.1007/BF01386390>.
- Dong, L., Li, J., Xia, W., Yuan, Q., 2021. Double ant colony algorithm based on dynamic feedback for energy-saving route planning for ships. *Soft. Comput.* 25 (7), 5021–5035. <https://doi.org/10.1007/s00500-021-05683-8>.
- Du, W., Li, Y., Zhang, G., Wang, C., Zhu, B., Qiao, J., 2022. Ship weather routing optimization based on improved fractional order particle swarm optimization. *Ocean Eng.* 248. <https://doi.org/10.1016/j.oceaneng.2022.110680>.
- Dui, H., Zheng, X., Wu, S., 2021. Resilience analysis of maritime transportation systems based on importance measures. *Reliab. Eng. Syst. Saf.* 209. <https://doi.org/10.1016/j.res.2021.107461>.
- Filipiak, D., Węcel, K., Stróżyna, M., Michalak, M., Abramowicz, W., 2020. Extracting Maritime Traffic Networks from AIS Data Using Evolutionary Algorithm. *Bus. Inf. Syst. Eng.* 62 (5), 435–450. <https://doi.org/10.1007/s12599-020-00661-0>.
- Funk, D., 2017. Analysis of the global maritime transportation system and its resilience. Naval Postgraduate School Monterey United States. <https://apps.dtic.mil/sti/citations/AD1046371>.
- Gao, M., Kang, Z., Zhang, A., Liu, J., Zhao, F., 2022. MASS autonomous navigation system based on AIS big data with dueling deep Q networks prioritized replay reinforcement learning. *Ocean Eng.* 249. <https://doi.org/10.1016/j.oceaneng.2022.110834>.
- Grifoll, M., Borén, C., Castells-Sanabra, M., 2022. A comprehensive ship weather routing system using CMEMS products and A* algorithm. *Ocean Eng.* 255, 111427. <https://doi.org/10.1016/j.oceaneng.2022.111427>.
- Hagiwara, H. (1989). Weather Routing of (Sail-Assisted) Motor Vessels. Ph.D. Thesis, Delft University of Technology, Delft.
- He, H., Mansuy, M., Verwilligen, J., Delefortrie, G., Lataire, E., 2024. Global path planning for inland vessels based on fast marching algorithm. *Ocean Eng.* 312. <https://doi.org/10.1016/j.oceaneng.2024.119172>.
- Hart, P., Nilsson, N., Raphael, B., 1968. A Formal Basis for the Heuristic Determination of Minimum Cost Paths. *IEEE Trans. Syst. Sci. Cybernet.* 4 (2), 100–107. <https://doi.org/10.1109/TSSC.1968.300136>.
- Huang, L., Zhang, Z., Wen, Y., Zhu, M., & Huang, Y. (2021). Stopping behavior recognition and classification of ship.
- James, R.W. (1957). Application of Wave Forecasts to Marine Navigation.
- Ji, Y., Qi, L., Balling, R., 2022. A dynamic adaptive grating algorithm for AIS-based ship trajectory compression. *J. Navigation* 75 (1), 213–229. <https://doi.org/10.1017/S0373463321000692>.
- Jiang, M., Lu, J., Qu, Z., Yang, Z., 2021. Port vulnerability assessment from a supply Chain perspective. *Ocean Coast. Manag.* 213. <https://doi.org/10.1016/j.ocecoaman.2021.105851>.
- Kaklis, D., Kontopoulos, I., Varlamis, I., Emiriz, I.Z., Varelas, T., 2024. Trajectory mining and routing: a cross-sectoral approach. *J. Marine Sci. Eng.* 12 (1). <https://doi.org/10.3390/jmse12010157>.
- Killick, R., Fearnhead, P., Eckley, I.A., 2012. Optimal detection of changepoints with a linear computational cost. *J. Am. Stat. Assoc.* 107, 1590–1598.
- Lehtola, V., Montewka, J., Goerlandt, F., Guinness, R., Lensu, M., 2019. Finding safe and efficient shipping routes in ice-covered waters: A framework and a model. *Cold Reg. Sci. Technol.* 165. <https://doi.org/10.1016/j.coldregions.2019.102795>.
- Liu, C., Ma, Y., Cao, C., Yan, X., 2023a. Ship route planning in the pirate area via hybrid probabilistic roadmap algorithm within the context of the Maritime Silk Road. *Ocean Coast. Manag.* 238. <https://doi.org/10.1016/j.ocecoaman.2023.106585>.
- Liu, D., Rong, H., Guedes Soares, C., 2023b. Shipping route modelling of AIS maritime traffic data at the approach to ports. *Ocean Eng.* 289. <https://doi.org/10.1016/j.oceaneng.2023.115868>.
- Li, H., Lam, J.S.L., Yang, Z., Liu, J., Liu, R.W., Liang, M., Li, Y., 2022. Unsupervised hierarchical methodology of maritime traffic pattern extraction for knowledge discovery. *Transp. Res. Part C Emerging Technol.* 143, 103856. <https://doi.org/10.1016/j.trc.2022.103856>.
- Li, L., Wu, D., Huang, Y., Yuan, Z.-M., 2021. A path planning strategy unified with a COLREGS collision avoidance function based on deep reinforcement learning and artificial potential field. *Appl. Ocean Res.* 113. <https://doi.org/10.1016/j.apor.2021.102759>.
- Li, Y., Li, H., Zhang, C., Zhao, Y., Yang, Z., 2024. Incorporation of adaptive compression into a GPU parallel computing framework for analyzing large-scale vessel trajectories. *Transp. Res. Part C Emerging Technol.* 163. <https://doi.org/10.1016/j.trc.2024.104648>.
- Lin, X., Wang, S., Zhang, X., Hsieh, T.-H., Sun, Z., Xu, T., 2021. Near-Field Route Optimization-Supported Polar Ice Navigation via Maritime Radar Videos. *J. Adv. Transp.* 2021, 1–15. <https://doi.org/10.1155/2021/2798351>.
- Liu, C., Kulkarni, K., Suominen, M., Kujala, P., Musharraf, M., 2024. On the data-driven investigation of factors affecting the need for icebreaker assistance in ice-covered waters. *Cold Reg. Sci. Technol.* 221, 104173.
- Liu, L., Shibasaki, R., Zhang, Y., Kosuge, N., Zhang, M., Hu, Y., 2023c. Data-driven framework for extracting global maritime shipping networks by machine learning. *Ocean Eng.* 269. <https://doi.org/10.1016/j.oceaneng.2022.113494>.
- Ma, D., Ma, W., Jin, S., Ma, X., 2020. Method for simultaneously optimizing ship route and speed with emission control areas. *Ocean Eng.* 202, 107170. <https://doi.org/10.1016/j.oceaneng.2020.107170>.
- Ma, W., Lu, T., Ma, D., Wang, D., Qu, F., 2021. Ship route and speed multi-objective optimization considering weather conditions and emission control area regulations. *Marit. Policy Manag.* 48 (8), 1053–1068. <https://doi.org/10.1080/03088839.2020.1825853>.
- Meng, Q., Wang, T., 2011. A scenario-based dynamic programming model for multi-period liner ship fleet planning. *Transport. Res. Part E: Logistics Transport. Rev.* 47 (4), 401–413. <https://doi.org/10.1016/j.trre.2010.12.005>.
- Moradi, M.H., Brutsche, M., Wenig, M., Wagner, U., Koch, T., 2022. Marine route optimization using reinforcement learning approach to reduce fuel consumption and consequently minimize CO2 emissions. *Ocean Eng.* 259, 111882. <https://doi.org/10.1016/j.oceaneng.2022.111882>.
- Naus, K., 2019. Drafting Route Plan Templates for Ships on the Basis of AIS Historical Data. *J. Navig.* 73 (3), 726–745. <https://doi.org/10.1017/s0373463319000948>.
- Papageorgiou, D.J., Cheon, M.-S., Nemhauser, G., Sokol, J., 2015. Approximate Dynamic Programming for a Class of Long-Horizon Maritime Inventory Routing Problems. *Transp. Sci.* 49 (4), 870–885. <https://doi.org/10.1287/trsc.2014.0542>.
- Park, J., Kim, N., 2015. Two-phase approach to optimal weather routing using geometric programming. *J. Mar. Sci. Technol.* 20 (4), 679–688. <https://doi.org/10.1007/s00773-015-0321-6>.
- Pennino, S., Gaglione, S., Innac, A., Piscopo, V., Scamardella, A., 2020. Development of a new ship adaptive weather routing model based on seakeeping analysis and optimization. *J. Marine Sci. Eng.* 8 (4), 270. <https://doi.org/10.3390/jmse8040270>.
- Poulsen, R.T., Viktoria, M., Varve, H., Rasmussen, H.B., von Knorring, H., 2022. Energy efficiency in ship operations - Exploring voyage decisions and decision-makers. *Transp. Res. Part D: Transp. Environ.* 102, 103120. <https://doi.org/10.1016/j.trd.2021.103120>.
- Shin, Y.W., Abebe, M., Noh, Y., Lee, S., Lee, I., Kim, D., Bae, J., Kim, K.C., 2020. Near-optimal weather routing by using improved A* Algorithm. *Appl. Sci.* 10 (17), 6010. <https://doi.org/10.3390/app10176010>.
- Szlapczynska, J., Smierzchalski, R., 2007. Adopted isochrone method improving ship safety in weather routing with evolutionary approach. *Int. J. Reliab. Qual. Saf. Eng.* 14 (06), 635–645. <https://doi.org/10.1142/S0218539307002842>.
- Tang, C., Wang, H., Zhao, J., Tang, Y., Yan, H., Xiao, Y., 2021. A method for compressing AIS trajectory data based on the adaptive-threshold Douglas-Peucker algorithm. *Ocean Engineering* 232. <https://doi.org/10.1016/j.oceaneng.2021.109041>.
- Tran, T.T., Browne, T., Musharraf, M., Veitch, B., 2023. Pathfinding and optimization for vessels in ice: A literature review. *Cold Reg. Sci. Technol.* 211. <https://doi.org/10.1016/j.coldregions.2023.103876>.
- Truong, C., Oudre, L., Vayatis, N., 2020. Selective review of offline change point detection methods. *Signal Process.* 167. <https://doi.org/10.1016/j.sigpro.2019.107299>.

- Tsou, M.-C., Hsueh, C.-K., 2010. The Study of Ship Collision Avoidance Route Planning by Ant Colony Algorithm. *J. Mar. Sci. Technol.* 18 (5). <https://doi.org/10.51400/2709-6998.1929>.
- Turna, İ., 2023. A Fuzzy Bayesian approach for 'Appraisal' of ship voyage plans. *Ships Offshore Struct.* 18 (6), 859–866. <https://doi.org/10.1080/17445302.2022.2077279>.
- UNCTD. (2022). Review of maritime transport 202. U. N. C. o. T. a. Development. https://unctad.org/system/files/official-document/rmt2022_en.pdf.
- Varlamis, I., Kontopoulos, I., Tserpes, K., Etemad, M., Soares, A., Matwin, S., 2020. Building navigation networks from multi-vessel trajectory data. *GeoInformatica* 25 (1), 69–97. <https://doi.org/10.1007/s10707-020-00421-y>.
- Wang, H., Lang, X., Mao, W., 2021. Voyage optimization combining genetic algorithm and dynamic programming for fuel/emissions reduction. *Transp. Res. Part D: Transp. Environ.* 90. <https://doi.org/10.1016/j.trd.2020.102670>.
- Wang, H., Lang, X., Mao, W., Zhang, D., Storhaug, G., 2020a. Effectiveness of 2D optimization algorithms considering voluntary speed reduction under uncertain metocean conditions. *Ocean Eng.* 200, 107063. <https://doi.org/10.1016/j.oceaneng.2020.107063>.
- Wang, H., Mao, W., Eriksson, L., 2019. A Three-Dimensional Dijkstra's algorithm for multi-objective ship voyage optimization. *Ocean Eng.* 186. <https://doi.org/10.1016/j.oceaneng.2019.106131>.
- Wang, K., Guo, X., Zhao, J., Ma, R., Huang, L., Tian, F., Dong, S., Zhang, P., Liu, C., Wang, Z., 2022a. An integrated collaborative decision-making method for optimizing energy consumption of sail-assisted ships towards low-carbon shipping. *Ocean Eng.* 266, 112810. <https://doi.org/10.1016/j.oceaneng.2022.112810>.
- Wang, L., Zhang, Z., Zhu, Q., Ma, S., 2020b. Ship Route Planning Based on Double-Cycling Genetic Algorithm Considering Ship Maneuverability Constraint. *IEEE Access* 8, 190746–190759. <https://doi.org/10.1109/access.2020.3031739>.
- Wang, X., Liu, Z., Yan, R., Wang, H., Zhang, M., 2022b. Quantitative analysis of the impact of COVID-19 on ship visiting behaviors to ports- A framework and a case study. *Ocean Coast Manag* 230, 106377. <https://doi.org/10.1016/j.ocecoaman.2022.106377>.
- Wei, Q., Liu, Y., Dong, Y., Li, T., Li, W., 2023. A digital twin framework for real-time ship routing considering decarbonization regulatory compliance. *Ocean Eng.* 278. <https://doi.org/10.1016/j.oceaneng.2023.114407>.
- Wei, Z., Xie, X., Zhang, X., 2020. AIS trajectory simplification algorithm considering ship behaviours. *Ocean Eng.* 216. <https://doi.org/10.1016/j.oceaneng.2020.108086>.
- Xin, J., Zhong, J., Yang, F., Cui, Y., Sheng, J., 2019. An Improved Genetic Algorithm for Path-Planning of Unmanned Surface Vehicle. *Sensors* 19 (11), 2640. <https://doi.org/10.3390/s19112640>.
- Xu, M., Pan, Q., Muscoloni, A., Xia, H., Cannistraci, C.V., 2020. Modular gateway-ness connectivity and structural core organization in maritime network science. *Nat Commun* 11 (1), 2849. <https://doi.org/10.1038/s41467-020-16619-5>.
- Xu, D., Yang, J., Zhou, X., Xu, H., 2024. Hybrid path planning method for USV using bidirectional A* and improved DWA considering the manoeuvrability and COLREGs. *Ocean Eng.* 298. <https://doi.org/10.1016/j.oceaneng.2024.117210>.
- Xin, X., Liu, K., Li, H., Yang, Z., 2024. Maritime traffic partitioning: An adaptive semi-supervised spectral regularization approach for leveraging multi-graph evolutionary traffic interactions. *Transp. Res. Part C Emerging Technol.* 164. <https://doi.org/10.1016/j.trc.2024.104670>.
- Yan, R., Mo, H., Yang, D., Wang, S., 2022. Development of denoising and compression algorithms for AIS-based vessel trajectories. *Ocean Eng.* 252. <https://doi.org/10.1016/j.oceaneng.2022.111207>.
- Yan, Z., Xiao, Y., Cheng, L., He, R., Ruan, X., Zhou, X., Bin, R., 2020. Exploring AIS data for intelligent maritime routes extraction. *Appl. Ocean Res.* 101, 102271. <https://doi.org/10.1016/j.apor.2020.102271>.
- Yin, Z., Yang, D., Bai, X., 2022. Vessel destination prediction: A stacking approach. *Transp. Res. Part C Emerging Technol.* 145, 103951. <https://doi.org/10.1016/j.trc.2022.103951>.
- Zhang, C., Bin, J., Wang, W., Peng, X., Wang, R., Halldearn, R., Liu, Z., 2020. AIS data driven general vessel destination prediction: A random forest based approach. *Transp. Res. Part C Emerging Technol.* 118, 102729. <https://doi.org/10.1016/j.trc.2020.102729>.
- Zhang, C., Zhang, D., Zhang, M., Mao, W., 2019. Data-driven ship energy efficiency analysis and optimization model for route planning in ice-covered Arctic waters. *Ocean Eng.* 186. <https://doi.org/10.1016/j.oceaneng.2019.05.053>.
- Zhang, C., Zhang, D., Zhang, M., Zhang, J., Mao, W., 2022. A three-dimensional ant colony algorithm for multi-objective ice routing of a ship in the Arctic area. *Ocean Eng.* 266, 113241. <https://doi.org/10.1016/j.oceaneng.2022.113241>.
- Zhang, G., Wang, H., Zhao, W., Guan, Z., Li, P., 2021a. Application of Improved Multi-Objective Ant Colony Optimization Algorithm in Ship Weather Routing. *J. Ocean Univ. China* 20 (1), 45–55. <https://doi.org/10.1007/s11802-021-4436-6>.
- Zhang, M., Montewka, J., Manderbacka, T., Kujala, P., Hirdaris, S., 2021b. A big data analytics method for the evaluation of ship-ship collision risk reflecting hydrometeorological conditions. *Reliab. Eng. Syst. Saf.* 213, 107674. <https://doi.org/10.1016/j.ress.2021.107674>.
- Zhang, M., Taimuri, G., Zhang, J., Zhang, D., Yan, X., Kujala, P., Hirdaris, S., 2025. Systems driven intelligent decision support methods for ship collision and grounding prevention: Present status, possible solutions, and challenges. *Reliability Engineering & System Safety* 253, 110489.
- Zhang, M., Tsoulakos, N., Kujala, P., Hirdaris, S., 2024. A deep learning method for the prediction of ship fuel consumption in real operational conditions. *Eng. Appl. Artif. Intel.* 130, 107425.
- Zhang, S., Shi, G., Liu, Z., Zhao, Z., Wu, Z., 2018. Data-driven based automatic maritime routing from massive AIS trajectories in the face of disparity. *Ocean Eng.* 155, 240–250. <https://doi.org/10.1016/j.oceaneng.2018.02.060>.
- Zhao, L., Shi, G., 2019. A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition. *Ocean Eng.* 172, 456–467. <https://doi.org/10.1016/j.oceaneng.2018.12.019>.
- Zhao, W., Wang, Y., Zhang, Z., Wang, H., 2021. Multicriteria ship route planning method based on improved particle swarm optimization–genetic algorithm. *J. Marine Sci. Eng.* 9 (4). <https://doi.org/10.3390/jmse9040357>.
- Zhou, Y., Daamen, W., Vellinga, T., Hoogendoorn, S., 2019. Review of maritime traffic models from vessel behavior modeling perspective. *Transp. Res. Part C Emerging Technol.* 105, 323–345. <https://doi.org/10.1016/j.trc.2019.06.004>.