Xin, X, Liu, K, Loughney, S, Wang, J, Li, H, Ekere, N and Yang, Z

Multi-scale collision risk estimation for maritime traffic in complex port waters

Reliability Engineering and System Safety 240 (2023) 109554

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Multi-scale collision risk estimation for maritime traffic in complex port waters

Xuri Xin [a,b], Kezhong Liu [a,c,d], Sean Loughney [b], Jin Wang [b], Huanhuan Li [b], Nduka Ekere [b], Zaili Yang [b,*]

[a] School of Navigation, Wuhan University of Technology, Wuhan, Hubei, China
[b] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, UK
[c] Hubei Key Laboratory of Inland Shipping Technology, Wuhan, Hubei, China
[d] National Engineering Research Center for Water Transport Safety, Wuhan University of Technology, Wuhan, Hubei, China

ARTICLE INFO

Keywords:
Maritime safety
Intelligent maritime surveillance
Multi-scale collision risk
Network theory
Graph clustering

ABSTRACT

Ship collision risk estimation is an essential component of intelligent maritime surveillance systems. Traditional risk estimation approaches, which can only analyze traffic risk in one specific scale, reveal a significant challenge in quantifying the collision risk of a traffic scenario from different spatial scales. This is detrimental to understanding the traffic situations and supporting effective anti-collision decision-making, particularly as maritime traffic complexity grows and autonomous ships emerge. In this study, a systematic multi-scale collision risk estimation approach is newly developed to capture traffic conflict patterns under different spatial scales. It extends the application of the complex network theory and a node deletion method to quantify the interactions and dependencies among multiple ships within encounter scenarios, enabling collision risk to be evaluated at any spatial scale. Meanwhile, an advanced graph-based clustering framework is introduced to search for the optimal spatial scales for risk evaluation. Extensive numerical experiments based on AIS data in Ningbo_Zhoushan Port are implemented to evaluate the model performance. Experimental results reveal that the proposed approach can strengthen maritime situational awareness, identify high-risk areas and support strategic maritime safety management. This work therefore sheds light on improving the intelligent levels of maritime surveillance and promoting maritime traffic automation.

1. Introduction

Maritime transportation management has been gaining considerable concern due to its paramount role in reducing accidents, improving economy, and protecting the ocean environment. Advanced technologies and policies related to perception, communication, digitization, and automation have accelerated the revolutionization of the maritime transport industry. Particularly, various modern and intelligent technologies and systems such as Internet of Things (IoT), cloud computing technologies, Automatic Identification System (AIS), and navigation aids and decision support systems have been deployed and incorporated into maritime transportation surveillance [1–3]. They present great potential to aid maritime authorities in proactive maritime traffic monitoring and surveillance. Meanwhile, large investments are conducted in maritime industrial projects to digitize maritime operational platforms, reduce manning requirements, and implement autonomous technology [4,5]. Following the promising development trend, the current maritime industry is evolving from conventional mechanical systems to digital systems with autonomous modules moving towards a reality [6,7]. Accordingly, further requirements on new situational awareness techniques are inevitable and increasingly demanded to promote maritime system intelligence and ship navigation automation.

The quantitative collision risk estimation of a maritime traffic situation is indispensable to Maritime Situational Awareness (MSA). It is of great benefit to the safety and efficiency enhancement of maritime transportation, such as providing early collision warnings [8,9], facilitating route planning [1], and evaluating risks in the nautical navigational environment [10]. However, economic globalization, the considerable growth in traffic demand, and the emergence of autonomous ships have incurred more complicated traffic situations, involving dynamic traffic movements, uneven traffic spatiotemporal distribution, restricted water topography and multiple dependent conflicts,

* Corresponding author.
E-mail address: Z.Yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.ress.2023.109554
Received 1 March 2023; Received in revised form 3 June 2023; Accepted 11 August 2023
Available online 12 August 2023

particularly in complex port waters [11,12]. The increasingly complex traffic scenarios, where many sophisticated traffic behavior characteristics are incorporated together, raise a new challenge to the applicability and effectiveness of the traditional risk estimation methods in the literature. Notably, the accuracy and reliability of a risk estimation model are key determinants for developing an autonomous shipping scheme. More specifically, autonomous systems generally consist of perception and control modules, where the former acts as a prerequisite to the decision-making design [13]. Therefore, new risk estimation models are urgently needed to handle the ever-growing complexity of traffic scenarios and to improve the intelligent levels of autonomous navigation systems.

In practice, maritime collision risk estimation in a busy water area exhibits significantly distinct properties in different spatial scales. From a micro spatial scale, the collision risk of ship pairs can be accurately measured, and the local conflict patterns are revealed with fine granularity. However, the characterization of global multi-conflict interactions on this scale remains a difficulty. On the contrary, the global collision risk can be easily recognized from a macro spatial scale, whereas the critical details of local conflict patterns are often ignored. Intuitively, an appropriate way to solve this issue is to incorporate the multi-scale patterns into risk evaluation modeling to interpret the traffic situation comprehensively and accurately. Unfortunately, due to the high complexity of this type of research, most existing studies [14,15] established collision risk models from a single scale. There has not been any systematic approach incorporating the multi-scale traffic properties into the maritime collision risk estimation in the literature. These existing works consequently encounter the difficulties in capturing the traffic conflict patterns under different granularity and in offering a complete comprehension for a traffic situation.

This study aims to develop a multi-scale collision risk estimation approach to achieve intelligent MSA in complex waters. It requires full traffic situation interpretability by extracting the traffic conflict patterns under different spatial scales and incorporating multiple dependent interrelationships relating to the dynamic co-behavior of multiple participating vessels. To address these issues, a holistic framework involving a set of designed technical models is built to reveal the actual traffic patterns under different spatial scopes and scales. In the framework, the near-miss collision risk measurement of ship pairs integrates the spatiotemporal dynamics of ship movements and restricted water geography by extending the classical Closest Point of Approach (CPA) based approach. The complex network theory is applied to support the quantification of spatiotemporal dependencies and interactions of multiple ships. Furthermore, the constructed framework integrates a graph clustering algorithm to adaptively partition the entire ship traffic into the optimal scales in terms of the spatiotemporal interrelationships among ships. Only by doing so, the traffic interaction analyses under different scales can be realised. The successful combination of these techniques will facilitate a better understanding of maritime traffic situations and promote maritime safety management in the context of developing intelligent and automated maritime traffic.

The new contributions of this paper are summarized as follows.

(1) A multi-scale collision risk estimation framework is proposed for the first time by synergizing a sequence of modeling techniques. Different from the traditional models that process collision risk in a single scale, it can capture the conflict patterns under different spatiotemporal granularity.
(2) An improved CPA-based method is developed to precisely quantify the collision risk of ship pairs in complex waters. It incorporates the influence of ship motion dynamics and restricted water topography on collision risk measurement, enabling it to be adaptive to various encountering scenarios in complex traffic waters.
(3) This study develops a regional/global collision risk evaluation model to characterize the topological characteristics associated with the interaction structures of entire ship traffic in a complex water. It also pioneers the application of a node deletion method in revealing the aggregation effect of multi-ship risk interactions on the entire traffic situation.
(4) To extract the traffic conflict patterns under different spatial scales, a competitive graph clustering technique is embedded to perform maritime traffic partition. It accounts for the spatiotemporal dependencies among multiple ships, making it desirable to adaptively select the optimal scopes for risk evaluation.

The remaining parts of this paper are organized as follows. Section 2 reviews the literature associated with maritime collision risk evaluation and estimation. Section 3 describes the new methodology and elaborates on the details of the multi-scale collision risk estimation approach. In Section 4, the application implementation and performance of the methodology are illustrated and discussed. Conclusions and future directions are outlined in Section 5.

2. Literature review

Maritime collision risk estimation is an integral component of safety management. It makes significant contributions to enhancing maritime navigation safety and avoiding potential collisions. Thanks to the accessibility of a large amount of AIS-based trajectory data and the improvement in data quality, the practical applications of AIS data in navigation safety-relevant research have drawn extensive concerns from academic circles, as reviewed in [16,17]. Similarly, many efforts have been placed on constructing highly reliable and scalable risk evaluation models (e.g., a probabilistic conflict detection model [18], a time-varying collision risk model [19], and an improved rule-aware time-varying collision risk measurement model [20]) and developing new risk measure concepts (e.g., collision candidate [21] and near-miss [22]). A comprehensive survey and overview are documented in [14, 15]. Broadly speaking, the relevant works fall into two categories: use of a micro-level approach and a macro-level modeling approach.

There is an abundance of works related to micro-level collision risk analysis and evaluation. Dependent on the applied models, they are categorized into ship domain-based and CPA-based methods. The former characterizes the collision risk based on the violation or overlapping levels of safety zones of encountering ships. Various ship domains have been proposed to identify the relationships between the ship attributes and the spatiotemporal proximity risk (e.g., see the review in [23]). Advanced domain models can improve the risk evaluation accuracy but the high model complexity impedes their practical usefulness in real-time when considering the computational overhead [6]. Additionally, the applications of a domain model for collision estimation require its combination with the trajectory prediction approaches because of its technical incompetence in motion prediction. The latter quantifies when and how close the encounter ships will be during the look-ahead horizon based on the assumption that ships are sailing linearly. It generally synthesizes the parameters, including Distance to CPA (DCPA), Time to CPA (TCPA) and other essential factors (e.g., relative bearing, Bow Crossing Range, environmental disruption, and ship maneuverability [24–28]), to calculate the Collision Risk Index (CRI). Currently, most commercial systems adopt this type of approach to detect potential collisions due to its simple implementation and reliable performance when vessels keep a linear motion [29]. However, its accuracy becomes questionable when a vessel changes its course or speed. Therefore, further improvements are necessary for its practical use by incorporating the spatiotemporal dynamics of vessel movements.

Compared with the micro-level risk modeling, there is much less literature on macro-level collision risk evaluation and measurement. Only a few research studies have built regional/global risk models by taking density complexity into consideration [30], incorporating the collision risk and contribution of each vessel [31], integrating the unpredictability and irregularity of maritime traffic time sequences

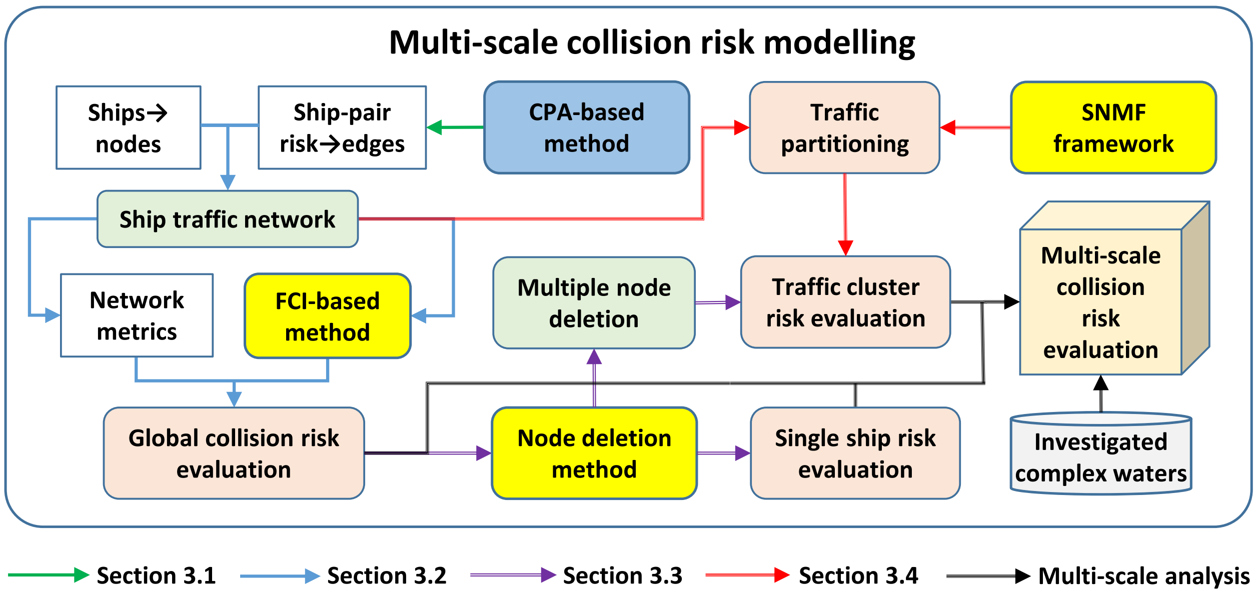

Fig. 1. Methodological framework of multi-scale collision risk estimation.

[32], and taking into account the evolutionary and structure properties of ship traffic networks [33]. Indeed, the main challenge for regional/global maritime risk analysis is on how to explicitly reveal the complexity of a scenario associated with the dependent conflict relations among multiple ships. In a heavy-traffic and complex water, the behaviours of ship traffic are spatially correlated with the structure of traffic conflicts. For instance, a single ship is not only influenced by its nearby conflicts, but also will be involved in the conflicts with far vessels as time goes by. Within this context, there has been a growing trend toward applying the complex network theory to unveil the topological properties of traffic interactions, especially in the air transportation field [34–37]. It has a rational and reliable performance in investigating the interrelations among different elements in a system and capturing the co-behavior features relating to the element interactions. In other words, the success of the complex network theory as an appropriate solution to describe the global traffic complexity stimulates this novel investigation of dynamic interactions of ship traffic in maritime transportation.

Although showing some attractiveness, the aforementioned methods can only process the collision risk in one specific spatial scale, which is inadequate to reveal the collision risk patterns under different spatial granularity. A systematic approach that can extract both the micro and macro spatial features enables the traffic situation to be described more precisely and completely. This is crucial to the intellectualization and automation of future maritime surveillance systems, especially considering its role in improving traffic pattern interpretability. Hence, much effort should be placed in analysing the correlation between the micro and macro collision risk to realize a proper combination of the multi-scale risk patterns. In the field of network analysis, many practices have shown that a node deletion method is a useful technique to characterize the relations between the whole system and its individual units. Notably, it has been successfully adopted to capture the crucial airports and ports in the whole traffic networks [38,39]. Inspired by these works, this study extends the application of the node deletion method to investigate the aggregation risk criticality of single/multiple ships to a regional/global traffic situation, thereby achieving a desired collision risk evaluation for any single/multiple ships in a busy water area of interest.

Another difficulty with multi-scale collision risk evaluation is extracting the optimal multi-ship clusters at different scales for risk assessment. Empirically, a gridding method is a commonly used tool to discretize the target maritime traffic zone into a grid-decomposed geographical space, where the size of each cell is usually predefined and unified. This method is adaptive to the batch analysis of historical traffic data by conducting the necessary traffic feature statistics within each cell (e.g., average speed, course change, and risk levels) to support maritime traffic visualization and collision avoidance [40–42]. Unfortunately, it cannot incorporate the complicated interactions relating to the real-time multiple dependent conflicts among ships when implementing the discretization. Generally, the spatial distribution of maritime traffic is unevenly spread over the water area, and the spatial dependencies among multiple conflicts may experience high dynamics over time. As a result, the issue concerning how to adaptively recognize the optimal traffic cluster scopes based on the real-time multi-ship dependencies becomes fundamental and requires much further investigation. Recently, the detection methods for encountering ship traffic clusters based on their spatial distance have attracted significant research interest [31,43,44]. What these studies have adopted is a density-based clustering technique, e.g., Density-Based Spatial Clustering of Applications with Noise (DBSCAN). This type of clustering technique focuses on the traffic density properties but reveals some drawbacks when incorporating other attribute-based interactions (e.g., collision risk) among ships. The graph clustering approach, which has been successfully used to capture congested road traffic regions [45–48], shows much attractiveness in integrating various interrelationships among the investigated objects during the clustering process. It is therefore adopted to partition the regional ship traffic into compact, scalable and interpretable traffic clusters.

In summary, the multi-scale collision risk qualification for ship traffic in complex waters is a very high-valued but complex work. The existing research remains challenging in identifying the correlation between the local and global collision risk, automatically capturing the optimal scales of risk analysis, and incorporating the complicated traffic characteristics in complex waters jointly. A system-level solution that makes effective synergies of various advanced techniques to perform multi-scale collision risk assessment is missing in the reported literature. Therefore, this study aims to develop a holistic framework to cope with these needs to allow the traffic patterns under different granularity to be extracted and evaluated. It brings new insights that have yet been revealed in the current literature from both theoretical and applied perspectives, hence making significant contributions to the formulation of operation services for intelligent maritime safety management and anti-collision solutions to autonomous ships.

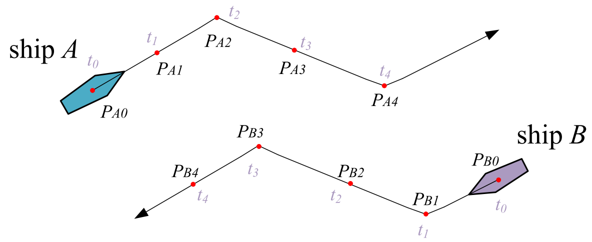

Fig. 2. An example of two encountering ships with dynamic movements.

3. Methodology: Multi-scale collision risk estimation

Fig. 1 presents a systematic framework of the proposed multi-scale collision risk estimation scheme. It involves a set of techniques which work holistically to explore the correlation between the collision risk under different types of spatial granularity. These techniques are effectively integrated with a series of interrelated steps, characterized by the following modules. Firstly, an improved CPA-based model incorporating the spatiotemporal dynamics of ship movements and restricted water topography is introduced to estimate the collision risk between ship pairs. Subsequently, a ship traffic network is established to evaluate the collision risk of the entire ship traffic in a given water area. It consists of nodes representing ships and links representing the collision risk of ship pairs which connect the nodes. In this stage of the research, five network indices from the complex network theory are used to quantify the regional traffic risk, which concerns traffic density, collision risk severity, and topological structure related to multi-ship interactions. In the meantime, a Fuzzy Clustering Iterative (FCI) method is applied to support a hierarchical and fine-grained assessment of multiple index synthesis. After these two steps, a node deletion method is utilized to examine the risk criticality of single or multiple ships to the traffic situation as a whole, to quantitatively estimate the ship traffic risk under any spatial scale. Additionally, the graph clustering model partitions the regional/global maritime traffic in terms of the spatiotemporal interrelationships among ships, and the collision risk assessment is conducted for the generated traffic clusters accordingly. This step addresses the problem of adaptive identification of the optimal spatial scales. By doing so, the multi-scale patterns of collision risk embedded into the complete maritime traffic situation can be extracted and evaluated, thereby achieving a comprehensive evaluation of the traffic scenarios. The technical details of relevant steps are highlighted in the following subsections.

3.1. Collision risk estimation of ship pairs

For the encounter situation analysis, the widely used CPA-based method is applied to evaluate whether the ship pairs have collision risk during the look-ahead horizon. Specifically speaking, this method calculates the DCPA and TCPA indicators relying on the hypothesis that the encountering ships will sail with a linear speed over a finite look-ahead horizon. However, the ships might have to take turning maneuvers in some cases due to various perturbations such as restricted waterway topography, environmental disturbances, and uncertain navigation intention, especially in complex port waters. Hence, it is necessary to describe ship pairs' relative spatiotemporal proximity relationship by incorporating their potential movement dynamics. In the meantime, it is crucial to consider the influence of intricate water topography on the measurement of collision risk. For instance, in complex waters, the presence of landmasses or islands may obstruct two spatially adjacent ships. In such scenarios, there is no collision risk between them.

Given the aforementioned research needs, an improved CPA-based model is employed to accurately quantify the collision risk between ships in complex waters. Firstly, a dynamic CPA calculation method is utilized to determine the actual DPCA and TPCA under ship motion dynamics. As illustrated in Fig. 2, the two ships involved in the encounter may change their course multiple times. Consequently, the dynamic trajectories of the ships can be represented by a sequence of waypoints, where the lines connecting consecutive waypoints depict the navigation route. The dynamic CPA-based model computes the CPAs between all pairs of successive waypoints. The minimum CPA between the two ships engaged in the encounter is then identified. The detailed explanation of the dynamic CPA computation is found in [49].

After obtaining the actual CPA values, it is necessary to determine whether the ships are obstructed by obstacles at the closest point of approach. To make this judgement, a sufficient number of sample points need to be evenly extracted from the line connecting the positions of the two ships at the closest approaching point. If at least one sample point falls into the unnavigable water areas, it indicates the presence of obstacles between them. Note that the navigable and unnavigable water areas in the investigated waters can be identified based on the grid-based methods [50]. To be specific, the entire water area can be divided into grids, in which the spatial density distribution of ship traffic in each grid can be further analysed based on historical AIS data [51,52]. If the obstacles are present between the ships, the collision risk between them is disregarded due to the constraint imposed by the water topography. Otherwise, the collision risk between ships can be assessed based on the obtained CPA values.

It should be noted that both DCPA and TCPA are fundamental to collision warning in maritime navigation. The former reveals the severity of a potential collision, whereas the latter reflects the time duration available for the collision resolution. In light of this, an exponential function that refers to work in [53] is adopted to synthesize the two indicators, as shown in Eq. (1):

$$CR_{ij} = \begin{cases} \left(\dfrac{\gamma_{DCPA} - DCPA_{ij}}{\gamma_{DCPA}}\right)^{1+\frac{TCPA_{ij}}{\gamma_{TCPA}}}, & \text{if } 0 \le DCPA_{ij} \le \gamma_{DCPA}, 0 \le TCPA_{ij} \le \gamma_{TCPA} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $DCPA_{ij}$ and $TCPA_{ij}$ represent the two indicators between ships i and j obtained by the dynamic CPA computation method, γ_{DCPA} and γ_{TCPA} denote the prescribed threshold values for collision detection, which are dependent on the application environment. Here, γ_{DCPA} is set to be two nautical miles in terms of the possible accepted safety distances given in [54,55]. Simultaneously, γ_{TCPA} is denoted as 15 min since this study performs collision detection at the medium-term time range with reference to the work in [6]. Overall, Eq. (1) conforms to the safety requirement in maritime transportation, i.e., an encounter scenario with smaller DCPA and TCPA values is more dangerous than the one with larger values, which effectively characterizes the ship-pairs' spatial and temporal proximity.

3.2. Regional/global collision risk evaluation

Once the proximity relationships of all ship pairs are measured, the ship traffic network can be constructed. A graph $G(V_N, E_L)$ is adopted to describe the ship traffic network, where V_N denotes N ship nodes connected by L links E_L. The ship pairs with collision risk larger than 0 are connected by an edge. Each edge weight is equal to the collision risk of the connected ship pairs. On this basis, the regional/global traffic collision risk can be evaluated from the perspective of complexity based on the complex network theory. It consists of two important components: one is the selection of network metrics, which requires the full characterization of a regional traffic situation; the other is the comprehensive evaluation of multiple metric measures, which concerns the adopted techniques and approaches with which the chosen metrics can be effectively combined to quantify the entire network collision risk. The relevant network metrics and index synthesis technique are elaborated

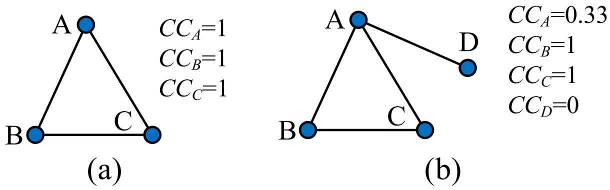

Fig. 3. An example of clustering coefficient calculation.

in the following sub-sections.

3.2.1. Network metrics

The complex network theory covers a variety of network metrics to characterize the structure-property of a network. In this study, five network metrics, including Number of Nodes (*NN*), Number of Edges (*NE*), Vertex Strength (*VS*), K-Shell Decomposition (*KS*) and Clustering Coefficient (*CC*), are used together to comprehensively reveal the global risk of a traffic scenario in a given water area. These metrics can capture distinct aspects of a ship traffic network, in which *NN* measures the traffic density in a given region, *NE* reflects the number of ship pairs that are at collision risk, *VS* quantifies the total collision risk, and *CC* and *KS* unveil the traffic network's topological characteristics related to the resolving difficulty of collisions. Their definitions are given as follows:

(1) *Number of Nodes* (*NN*) is treated as the basic feature of a network. It generally serves as a practical reference for the maritime operators to issue instructions. The higher the value of *NN*, the busier the traffic situation is.
(2) *Number of Edges* (*NE*) refers to the number of links connecting the node pairs. It reflects the number of ship pairs with the potential collision in the maritime traffic network. A larger *NE* corresponds to a riskier and more complex traffic situation, and vice versa.
(3) *Vertex Strength* (*VS*) represents the sum of edge weights correlated with one node. It integrates the characteristics of both the node degree and the associated edge weight, where the node degree refers to the number of nodes connecting with one specific node. Here the sum of all vertex strengths is used to unveil the total collision risk of a traffic scenario as shown in Eq. (2):

$$VS = \sum_{i=1}^{N} VS_i = \sum_{i=1}^{N} \sum_{j=1}^{D_i} w_{ij} \tag{2}$$

where N represents the number of nodes, w_{ij} denotes the edge weight between nodes i and j, and D_i is the number of adjacent nodes to node i. A high *VS* means that maritime traffic is more likely encountering a hazardous situation.
(4) *K-Shell Decomposition* (*KS*) is a typical technique that concerns the network structure. It partitions the network into several layers based on the coreness of the nodes. This metric works well in revealing how the nodes are grouped together and identifying the node's global important level. The nodes with dense connections are assigned with high KS_i values and the nodes in the same layer have identical KS_i indexes. The relevant details about k-shell calculation can be found in [56].This study adopts the maximum KS_i to reflect the difficulty level of conflict resolution caused by the traffic network topology. A larger *KS* means that lots of ships are spatially closer together with complicated interactions, and consequently, the surveillance controllers will encounter increased risk management pressure.
(5) *Clustering coefficient* (*CC*) quantifies the aggregation/clustering degree of the nodes in a graph. It can reveal how close the nodes' neighbours are to being a clique. The local *CC* of a node is denoted in Eq. (3):

$$CC_i = \frac{N_{\Delta}(i)}{d_i(d_i - 1)/2} \tag{3}$$

where $N_{\Delta}(i)$ denotes the real number of edges between the nodes that have connections with node i, and $d_i(d_i - 1)/2$ reflects the theoretical maximum number of edges between these nearby nodes. However, this metric cannot coincide with the basic principle of the global collision risk modeling, i.e., the increase of nodes or edges should not lead to the decline of the global risk. The example in Fig. 3 well justifies this. It is seen that the CC_A decreases when a new node and edge are embedded into the graph. Hence, an improved metric $CC_i^{'}$ that removes the denominator in Eq. (3) is developed. This new metric can quantify the complex interactions among the neighbours of one node while simultaneously meeting the global risk modeling principles. Furthermore, the sum of $CC_i^{'}$ is used to describe the global cross-conflict degree among ships, which is expressed using Eq. (4):

$$CC = \sum_{i=1}^{N} CC_i^{'} = \sum_{i=1}^{N} N_{\Delta}(i) \tag{4}$$

A larger *CC* is usually associated with more complex multiple dependent conflict-based interrelationships among ships and the corresponding conflicts within the traffic situation are harder to resolve.

3.2.2. Comprehensive evaluation of regional collision risk

FCI is in nature a typical multi-index evaluation method. It has gained the widespread concerns of scholars due to its success in various research fields [57–59]. Unlike other index integration techniques (e.g., entropy weight method, gray relational analysis, and analytic hierarchy process) that strongly rely on the evaluation standards and criteria, FCI focuses on the characteristics of the evaluated dataset and can deal with the fuzziness and uncertainty of the dataset [57,60]. Additionally, it allows the assessment objects to be evaluated in a hierarchical way. Hence, this study adopts it to support a comprehensive and fine-grained description of the regional collision risk.

In essence, FCI represents a class of data-driven techniques. It designs an iterative procedure to optimize the objective function continuously until it finds the optimal weight vector w, clustering center matrix S and membership matrix U [60]. The implementation details of this model can be found in Appendix A.

After obtaining the optimal w, U and S, the mapping relationships between the assessment samples and the different risk classes/grades can be calibrated. Assume that the regional collision risk is classified as c levels, and u_{kj} represents the membership degree of the jth regional traffic scenario belonging to the kth risk level, where $k = 1, 2, \ldots c$. Then two comprehensive indicators can be used to evaluate the regional collision risk, as shown by Eq. (5) and Eq. (6):

$$RCR_j^1 = \mathrm{argmax}_{\{k=1,2,\cdots,c\}} u_{kj} \tag{5}$$

$$RCR_j^2 = R_j \cdot w \tag{6}$$

where $R_j = (r_{1j}, r_{2j}, \ldots, r_{mj})$ is the normalized value of the adopted network metrics associated with the jth traffic scenario, and $w = (w_1, w_2, \ldots, w_m)$ denotes the influence weight of each metric in which m represents the number of network indices. The former indicator reveals the regional collision risk level/class based on each traffic scenario's maximum membership. By contrast, the latter reflects a single assessment value through the weighted sum of selected network indices.

3.3. Risk criticality of single/multiple ships

In addition to the collision risk evaluation for ship pairs and

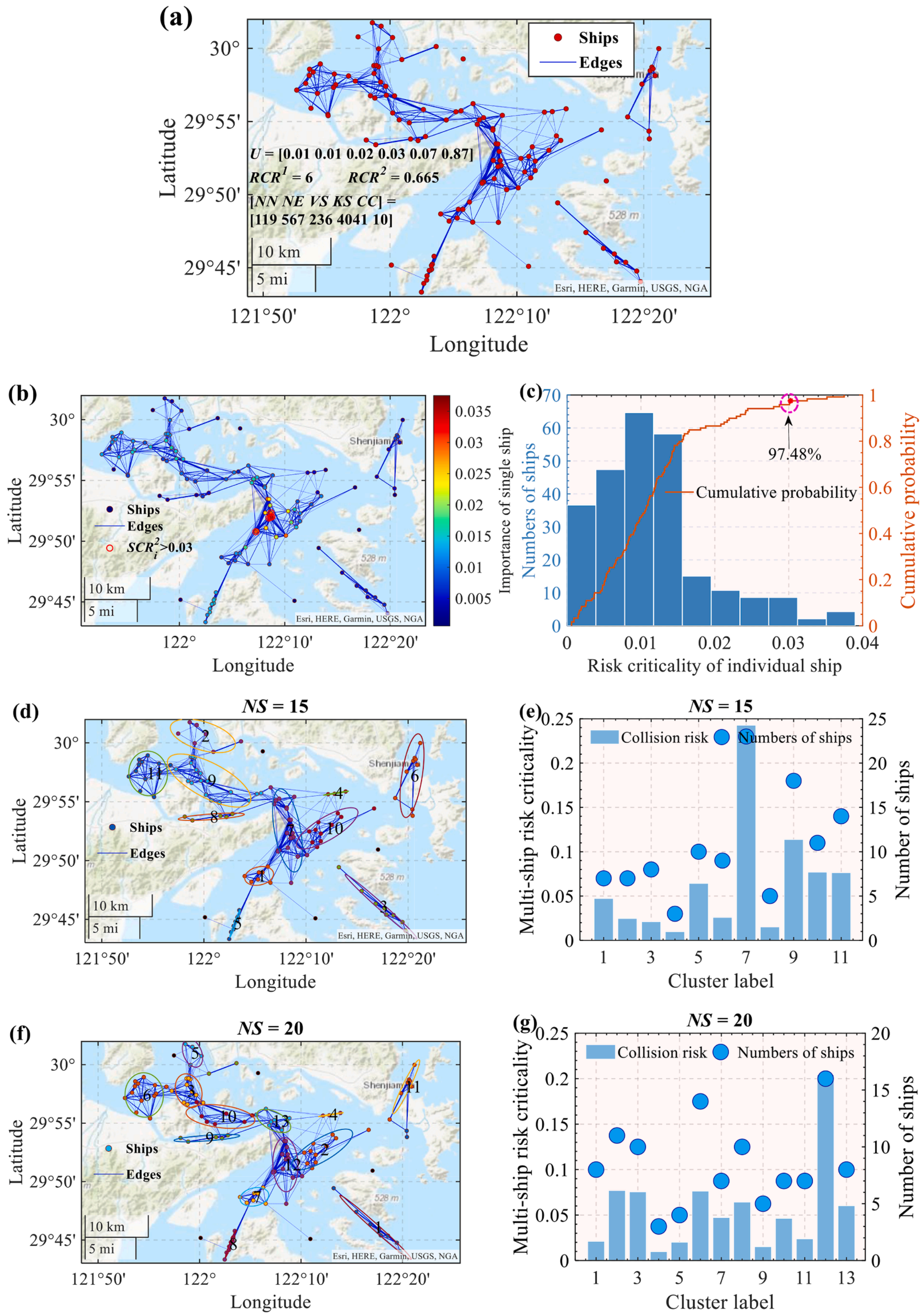

Fig. 4. Illustration of multi-scale collision risk of ship traffic at one time moment. (a) Visualization of ship traffic network; (b) visualization of risk criticality of any single ship to a regional traffic situation; (c) risk criticality distribution of single ships; (d)-(g) visualization of ship optimal traffic partition and collision risk of generated traffic clusters when $NS = 15$ and 20. Note that the generated traffic clusters with a number of ships less than three are not labelled.

regional/global ship traffic, the issue as to how the risk assessment under any spatial scale is realised is a remaining crucial question to answer. The node deletion method, as an effective system analytical tool, works well in quantifying the correlation between the whole network system and the single node [61]. Its essence is to measure each node's relative contribution by identifying the drops in the network performance resulting from the deletion of different nodes from the network. In this study, further extension and improvement are conducted for the node deletion method to explore both the risk contribution of any single and multiple dependent ships to a regional traffic situation. The contribution of the single/multiple ships to the regional collision risk can be interpreted as their risk criticality. The implementation process comprises two phases:

(1) *Risk criticality of single ship*: In this phase, the drops in the regional collision risk are calculated when each ship is removed from the

full traffic network separately. Suppose the regional collision risk associated with the full traffic network is given as RCR^2, and the regional collision risk when deleting the qth ship ($q = 1, 2, \dots ns$, where ns is the real-time number of ships in the entire network) from the full traffic network is computed as RCR_q^2. Then the margin of RCR_q^2 against RCR^2 can be calculated, denoted by $SCR_q^2 = RCR^2 - RCR_q^2$. Accordingly, the risk criticality of each individual ship to the regional traffic can be estimated based on the value of SCR_q^2. Additionally, the set of $SCR^2_{\{q\ =\ 1,\ 2,\ \dots,\ ns\}}$ can be ranked in a descending order, so that the ships with high-risk criticality values can be captured.

(2) *Risk criticality of multiple ships*: In this phase, the coupling risk contribution of multiple ships on a regional traffic situation is analysed. It investigates the aggregation risk criticality of multi-ships by removing them simultaneously from the full traffic network. It is noteworthy that this step should be combined with a maritime traffic partition approach since the latter helps detect multi-ship clusters with high spatiotemporal dependencies. The detailed procedure for maritime traffic partition is given in Section 3.4.

3.4. Regional maritime traffic partition

Maritime traffic partition is a crucial component for multi-scale collision risk analysis. It can recognize the optimal spatial scopes for risk evaluation, facilitating traffic pattern exploration under different granularity. In this study, a graph clustering technique called Symmetric Nonnegative Matrix Factorization (SNMF) is used for optimal ship traffic partition. It is a classical clustering algorithm used to address graph cut problems, facilitating the formation of clusters with high intra-cluster interactions and low inter-cluster interactions. Unlike other clustering algorithms like density-based and prototype-based clustering, SNMF focuses more on the interactions between pairs of samples rather than the attributes of each data sample. It takes a nonnegative similarity matrix as an input and approximates the similarity matrix with symmetric nonnegative low-rank matrices. According to a comprehensive survey in [62], SNMF can achieve higher accuracy and quality in addressing specific graph partition issues and demonstrates better performance than other graph clustering algorithms. These properties make it appealing for maritime traffic partition.

Many graph clustering approaches have consistent objective functions, which can be transformed into a trace maximization form as shown in Eq. (7):

$$\max_{H\geq 0, H^T H = I} \boldsymbol{Tr}\left(H^T W H\right) \tag{7}$$

where $W \in \mathbb{R}^{N\times N}$ denotes the similarity matrix; $H \in \mathbb{R}^{N\times NS}$ represents the clustering assignment matrix, satisfying $H \geq 0$ and $H^T H = I$; $\boldsymbol{Tr}$ is the trace of a matrix; and N and NS are the number of samples and clusters, respectively. Based on the constraints on H, Eq. (7) can be further mathematically equivalent to the following expressions:

$$\begin{aligned} &\min_{H\geq 0, H^T H = I} Tr\left(W^T W\right) - 2Tr\left(H^T W H\right) + Tr\left(H^T H\right) \\ \Leftrightarrow &\min_{H\geq 0, H^T H = I} \| W - HH^T \|^2 \end{aligned} \tag{8}$$

Since the optimization of Eq. (8) is an NP-hard issue considering the two constraints on H, SNMF seeks to relax the orthogonality constraint to make the formulation tractable. In [63], it has been verified that keeping the nonnegativity constraint can contribute to $H^T H \approx I$, which shows the feasibility and practicality of SNMF for an optimal graph partition.

To generate the traffic clusters with a good balance in size, the Normalized Cut objective function (Ncut) [64] is adopted to strike a balance between the sizes of different clusters. It minimizes the inter-cluster similarity but maximizes the intra-cluster similarity by using Eq. (9):

$$\mathrm{Ncut}(C_1, C_2, \cdots, C_{NS}) = \sum_{i=1}^{NS} \frac{cut(C_i, \overline{C}_i)}{cut(C_i, C)} \tag{9}$$

where C_i denotes the subset of the ith cluster, $\overline{C}_i$ represents the complement of C_i, C represents the entire data samples, $Cut(C_i, \overline{C}_i) = \sum_{u\in C_i, v\in \overline{C}_i} W_{uv}$, and W_{uv} is the similarity between samples u and v. Indeed, the minimization of Ncut is equivalent to replacing the W in Eq. (8) using a normalized similarity matrix $\widetilde{W} = D^{-1/2} W D^{-1/2}$. Accordingly, given the similarity matrix $\widetilde{W}$ and the desired number of clusters NS, the SNMF framework for maritime traffic partition can be given as follows:

$$\min_{H\geq 0} \| \widetilde{W} - HH^T \|^2 \tag{10}$$

where $\widetilde{W} \in \mathbb{R}_+^{N\times N}$ and $H \in \mathbb{R}_+^{N\times NS}$ are two matrices in which all the elements are enforced to be non-negative.

Note that each row in the lower rank matrix approximation H indicates the membership values of each sample belonging to different clusters. Hence, once obtaining the optimal H, one can identify the assignments of the clustered ships in terms of the largest element in each row in H. The detailed algorithmic steps for the traffic partition are depicted in Appendix B.

4. Case study: Implementation and results

To evaluate the effectiveness and feasibility of the proposed multi-scale collision risk methodology, the Ningbo-Zhoushan Port, as the world largest port in terms of throughput, is considered as the test site (see Fig. 4(a)). It presents one of the densest areas of maritime traffic in the world. More concretely, its restricted geographical waters, the high percentage of large-scale vessels, the dynamic ship motion behavior, and the uneven distribution of ship traffic expose it as a complicated and challenging scenario for maritime traffic risk analysis. These attributes jointly pose great challenges for maritime supervisors in effective MSA. Therefore, it is highly desirable to make use of this complex water area to validate the proposed methodology.

The AIS-based vessel trajectory data in the Ningbo-Zhoushan Port is deployed to constitute the input to maritime traffic analysis. It is a reliable source of information to be applied to various topics in the maritime traffic domain, including but not limited to vessel anomaly detection [65,66], marine traffic characteristics statistics [67–69], ship collision avoidance [70], vessel motion prediction [71,86], and trajectory behavior pattern extraction [72–74,85]. This study collected one month of AIS messages from 01/11/2018 to 30/11/2018, with the region under analysis bounded between latitudes 29°43′N-30°02′N and longitudes 121°52′E-122°22′E. Given that the AIS information errors are inevitable because of various technical issues, the AIS data pre-processing procedure presented in [75,76] is employed to reconstruct clean and accurate traffic trajectory. Additionally, a linear interpolation method proposed by [77] is adopted to capture the snapshot of the maritime traffic situation since the trajectory messages of different vessels are transmitted at varying frequencies. In this way, the reliable and complete information involved in a traffic scenario at any time moment can be extracted for traffic situation analysis and evaluation; that is, the AIS data pre-processing is the prerequisite of the real-time traffic situation analysis.

The rest of this section is organized as follows. Section 4.1 starts with a real application case to demonstrate how the proposed methodology is applied to evaluate the collision risk under different spatial scales. In Section 4.2, the collision risk distributions of regional traffic, single ships, and multiple ships are investigated, to assist in the identification and monitoring of critical high-risk traffic clusters. Section 4.3 describes the performance of a few statistics analyses of the traffic collision risk

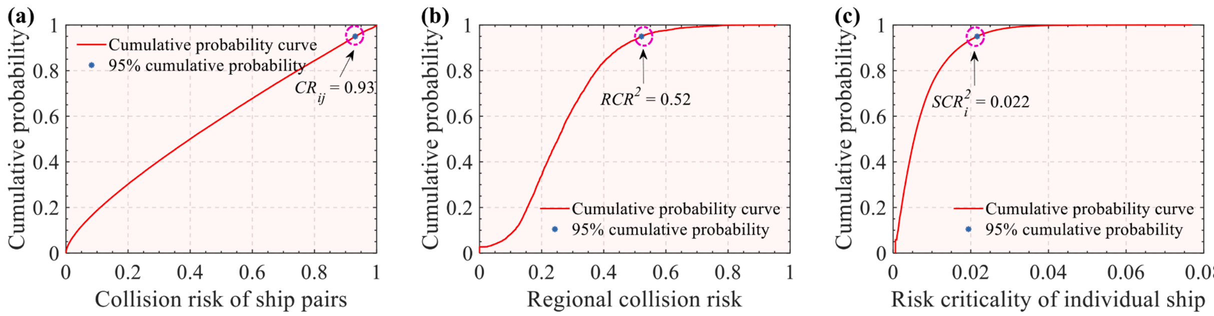

Fig. 5. (a) Cumulative probability distribution of collision risk of ship pairs; (b) cumulative probability distribution of regional collision risk; (c) cumulative probability distribution of single ship risk criticality.

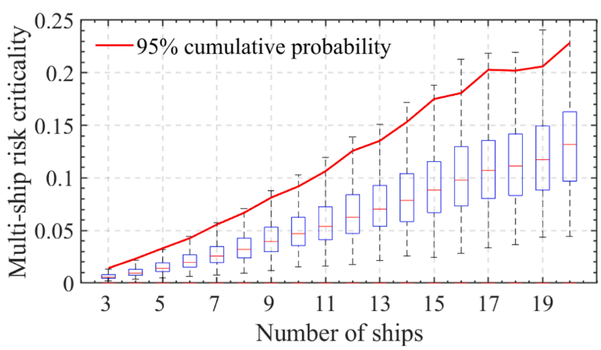

Fig. 6. Boxplot of risk criticality of traffic clusters with different numbers of ships.

evolution, to shed light on the future collision risk estimation and control. Furthermore, the model validation and comparison analysis are conducted in Section 4.4.

4.1. Application performance analysis

Fig. 4 provides the multi-scale collision risk analytical results for a specific traffic scenario within the research area. In Fig. 4(a), the visualization of the constructed traffic network is exhibited, where the red points stand for the spatial distribution of ship traffic, and the blue lines represent that the connected ship pairs have potential collision risk. It is found that there are complicated dependent interrelationships among ships. This phenomenon highlights the necessity of incorporating the topological indices (e.g., *KS* and *CC*) to describe the multi-ship interactions rather than merely concerning the ship-pair interactions when conducting a regional collision risk assessment. Based on the Equations in Section 3.2.1, the values of the adopted five network indicators for this scenario are 119, 567, 236, 4041, and 10, respectively. The regional collision risk is divided into six levels in terms of the FCI model, which are very low, low, slightly low, slightly high, high, and very high. The detailed training process for the FCI model is presented in Appendix C. The memberships belonging to different regional risk levels are then calculated as 0.01, 0.01, 0.02, 0.03, 0.07, and 0.87 in terms of Eq. (A.4) and the optimal weight vector (w) and class center matrix (U) in Appendix C. These membership results show a hierarchical description of the regional collision risk. The regional risk level and value (i.e., RCR^1 and RCR^2 in Eqs. (5) and (6)) are further obtained as 6 and 0.665, respectively, implying the complex situation of the analysed traffic scenario. These evaluation indicators work together to facilitate the interpretation of the regional/global traffic situations.

Fig. 4(b) displays the visualization of the single ship criticality to the regional collision risk. According to Fig. 4(b), the ships with higher criticality values can be easily captured based on their color indices. For instance, the ships with $SCR_i^2 > 0.03$ are highlighted with red circles. This can provide valuable guidelines for ship navigators to notice potential collision risks. Besides, the value distribution of single ship risk criticality is exhibited in Fig. 4(c). It is observed that the criticality of these ships is obviously heterogeneous and the key influential ships (The ships with $SCR_i^2 > 0.03$) account for a smaller percentage. Hence, the recognition of key influential ships can provide vital support in risk management from a global surveillance perspective. More specifically, the precise guidance and maneuvering instructions for these few critical ships can effectively aid to quickly mitigate the regional traffic complexity and consequently obtain the maximum regional collision risk reduction.

Furthermore, Fig. 4(d) and (f) illustrate visualization of ship optimal traffic partition when $NS = 15$ and $NS = 20$. In these figures, the ships in the same traffic cluster are spatially compact and have highly dependent conflict-related interrelationships. This indicates the good properties of the proposed traffic partition approach in searching for the optimal spatial scopes for risk evaluation. In practice, the traffic partition approach also aids to decrease the difficulty of situational awareness by decomposing the whole traffic situation into several compact and interpretable sub-clusters. Additionally, the aggregation risk criticality and the number of ships of each traffic cluster are presented in Fig. 4(e) and (g). The identified multi-ship aggregation effect of each cluster on the regional traffic risk can provide a practical reference for capturing the critical traffic clusters and resolving traffic conflicts in terms of the joint guidance of the multiple ships. Therefore, the combination of maritime traffic partition and multiple node deletion methods enables to extract and reveal the collision risk pattern under different granularity. Overall, the proposed multi-scale collision risk methodology supports a full comprehension of a specific navigation scenario.

4.2. Statistical analysis of traffic collision risk characteristics

It is important to note that the traffic clusters with larger numbers of ships generally correspond to higher collision risk. This is due to the fact that a larger traffic cluster usually has a more significant aggregation risk criticality to regional traffic. Therefore, the collision risk of traffic clusters should be compared in the same order, so that the key influential ships or traffic clusters with high risk can be more reasonably recognized and monitored. In light of this, the collision risk distributions of regional traffic, single, and multiple ships are statistically analysed separately. Fig. 5 illustrates the cumulative probability distributions of collision risk for ship pairs, regional traffic, and single ships. For example, the collision risk values corresponding to a 95% cumulative probability are chosen for a risk alert application. As a result, 0.93, 0.52, and 0.022 are the high-risk lines for the above three cases. Similarly, the collision risk criticality distributions of traffic clusters with different numbers of ships are displayed in Fig. 6. These analytical results offer a quantitative reference to trigger an earlier alert under different spatial granularity.

Built on the determined collision alert thresholds, Fig. 7 presents the

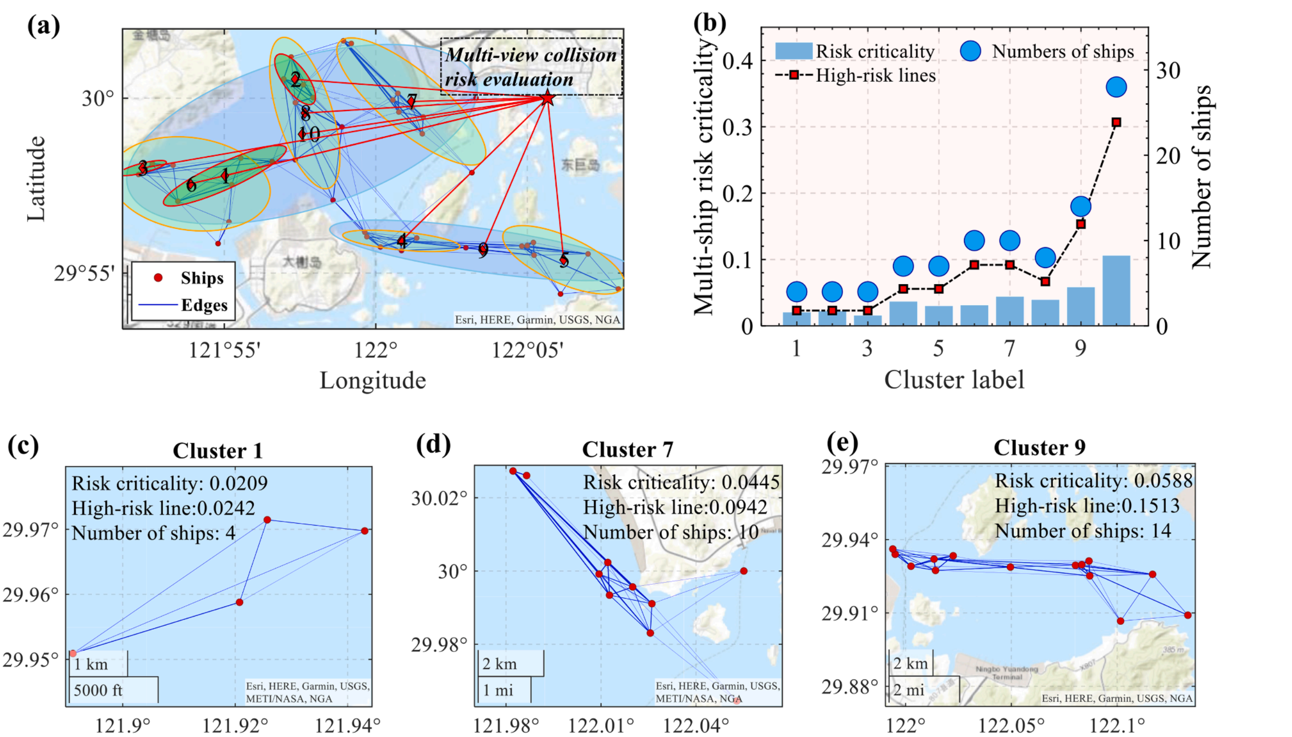

Fig. 7. Multi-view collision risk evaluation for a traffic scenario. (a) Visualization of multi-view traffic clusters; (b) risk criticality and number of ships of each traffic cluster; (c)–(e) visualization of some traffic clusters.

multi-view collision risk analysis results for a traffic scenario. In Fig. 7 (a), the multi-scale traffic patterns are revealed, including both small-scale (e.g., Cluster 3) and large-scale (e.g., Cluster 10) patterns. Fig. 7 (b) further exhibits the collision risk criticality and numbers of ships of different traffic clusters. It is seen that the risk criticality of Clusters 1 and 2 is close to the associated high-risk lines. Hence, much attention should be paid to these two clusters. In the meantime, Fig. 7(c)–(e) display the detailed information of Clusters 1, 7, and 9 to help better understand the generated clusters' multi-resolution features. The results reveal that the proposed methodology can proactively capture the high-risk areas under any spatial scale by integrating the determined collision alert thresholds.

4.3. Traffic collision risk evolution

A maritime traffic situation is inherently a dynamic evolving process; hence, multi-scale collision risk patterns commonly vary with time. To fully understand the evolutionary mechanism of the traffic situation, the time-dependent characteristics of collision risk are explored from three scales: regional, individual, and multiple ships.

Fig. 8 provides the evolutionary characteristics of regional collision risk. In Fig. 8(a) and (b), the transition probabilities between different regional collision risk levels are illustrated. Two insightful phenomena can be drawn. First, the cases that maintain the same risk levels (see the red bar in Fig. 8(a)) or transfer to their nearest risk levels (see the blue bar in Fig. 8(a)) over a short-term period almost account for 100%. Secondly, the jumping transitions, i.e., the transitions beyond 1 level, start to occupy a certain proportion when the evolving time lasts for 20 min (see the second bar in Fig. 8(b)). These results imply that the regional traffic situations evolve steadily over time. Meanwhile, the change degree of RCR^2 over short-term and long-term periods is shown in Fig. 8(c) and (d). According to these figures, the degree of change of RCR^2 grows linearly over a short-term period, while the growth rate of change degree gradually decreases over a long-term period. This is basically in line with the evolutionary features of RCR^1 (see the change degree of red bars in Fig. 8(a) and (b)). Additionally, Fig. 8(e) illustrates the life cycle of different regional collision risk levels. The life cycle refers to the existing duration of the current risk levels/classes. It is found when the regional collision risk is at higher risk levels, the "life cycle $\leq$ 5 min" will occupy a higher percentage; that is, a larger risk level corresponds to a shorter life cycle. This may be attributed to the fact that when faced with a persistent high-risk situation, maritime operators will take appropriate strategies to relieve the regional traffic complexity and consequently, the high-risk traffic situation will disappear soon. These evolutionary investigations provide a basis for future prediction of a high-level maritime traffic risk and offer insights into the design of maritime safety management strategies.

A similar evolutionary analysis is conducted for single ship collision risk. To investigate the transition features and life cycle of single ship collision risk, it is equally divided into six levels in terms of the cumulative probability distribution of individual collision risk. The corresponding statistical results are shown in Fig. 9. Compared to the regional collision risk evolution, two different findings are revealed. First, the individual collision risk evolves faster than the regional collision risk because the jumping transitions of risk levels occur over a short period (Compare Figs. 8(a) and 9(a)). Second, the medium risk levels have a shorter life cycle than the remaining risk levels since their life cycle with the duration less than or equal to 5 min occupies a higher percentage (see Fig. 9(c)). This may be because the slightly high (SH) and slightly low (SL) risk levels can transit to two sides while the very high (VH) and very low (VL) risk levels can only transit to one side, so that the formers are associated with higher dynamics and instability.

As for the evolutionary analysis of multiple ships, it is difficult to be performed since the generated traffic clusters will dynamically change over time rather than being time-invariant. To prove this, an inconsistent measure index in [78] is adopted to quantify the distance between two clustering results. Fig. 10 depicts the distance measure results between the traffic partitions with different time intervals. The numbers of traffic clusters applied for segmenting all traffic scenarios are 15 and 25, respectively. In theory, the distance index CT_{KM} of two identical traffic partition results is equal to the minus number of clusters, i.e. *-NS*. However, there is an apparent difference between the real CT_{KM} and

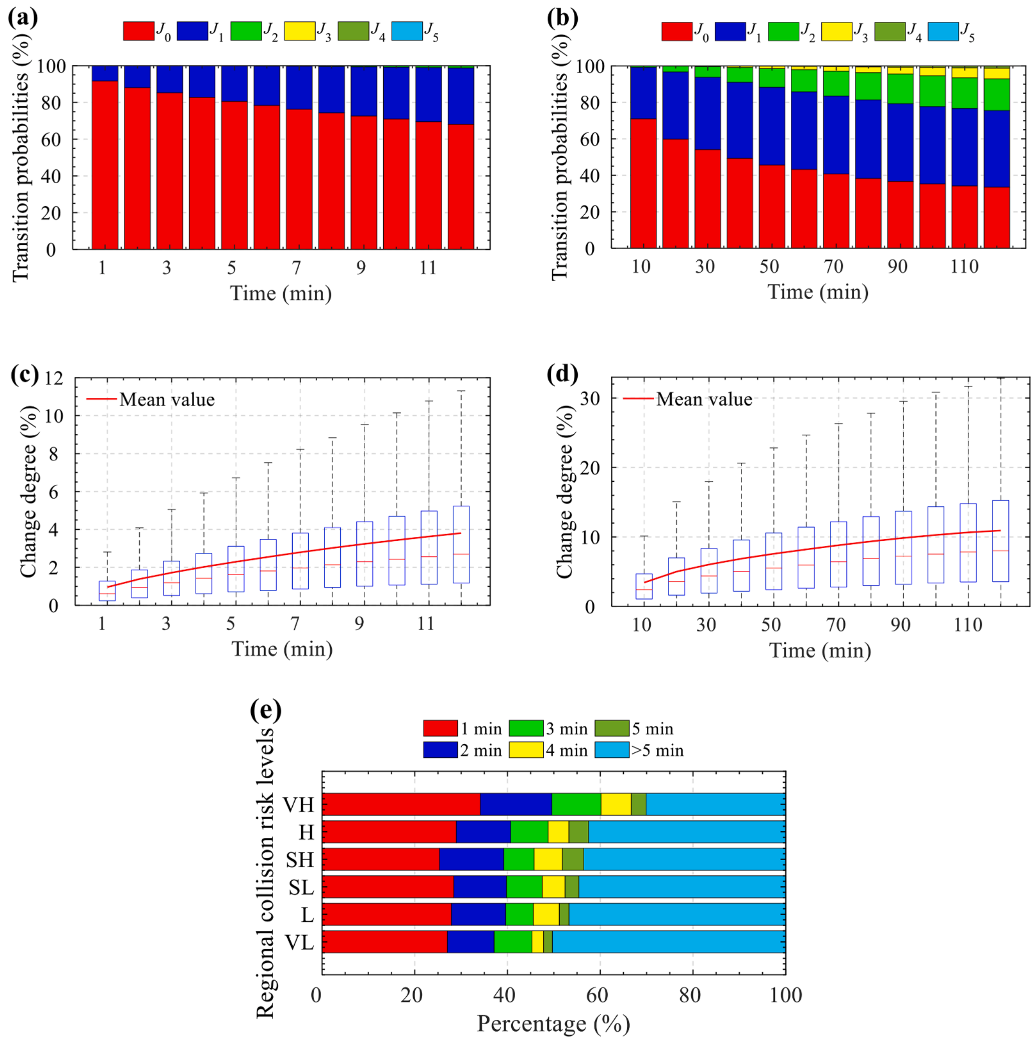

Fig. 8. Regional collision risk evolution. (a)-(b) Transition probabilities between different regional collision risk levels RCR1 over short-term and long-term periods, where J_0, J_1, …, J_5 represent the transitions with 1, 2, …, 5 levels; (c)-(d) change degree of RCR_2 over short-term and long-term periods; (e) life cycle of different regional collision risk levels, where 1 min, 2 min…, >5 min denote the life cycle with 1 min, 2 min…, >5 min.

ideal CT_{KM} (i.e., -15 in Fig. 10(a) and -25 in Fig. 10(b)), even for the CT_{KM} calculation of two partition results with a 1-minute interval. It reveals why conflict resolution involving multiple ships with spatio-temporal dependencies in highly dynamic waters is a complex task. It also indicates that the new traffic partition research incorporating the time-dependency nature of maritime traffic deserves further attention, which can obtain stable traffic clusters to support the continuous implementation of risk control strategies.

4.4. Model validation

The methodological validation is an indispensable part of any modeling approach since it confirms the confidence level of the results produced. In this study, the model robustness validation consists of two crucial blocks: one is the reliability analysis of the multi-scale collision risk evaluation, which needs to be conducted from the perspectives of ship pair, global traffic, single ship, and multiple ships, respectively; the other is the effectiveness test of the optimal maritime traffic partition, which concerns the model performance in terms of capturing the optimal spatial scales.

First, the effectiveness of the ship pair collision risk model in considering the influence of ship motion dynamics and water topography is analysed. A complicated two-ship encounter scenario is extracted from the historical AIS data, where the curves in Fig. 11(a) represent the trajectories of two encountering ships. It is obvious that the two ships have significant dynamic motion characteristics, and the presence of islands obstruct them. In Fig. 11(b), the *C-DCPA* and *E-DCPA* represent the DCPA values obtained by classic and dynamic CPA calculation models, respectively. Note that to obtain the waypoints of potential ship motion for *E-DCPA* calculation, the Douglas–Peucker (DP) algorithm is adopted to identify the turn points of ship trajectories and detailed implementation information about the algorithm is found in [79]. According to Fig. 11(b), the *C-DCPA* values fluctuate dramatically and experience several peaks and troughs, which may confuse ship navigators in capturing the actual collision dangers. In contrast, the *E-DCPA* values keep relatively stable and can help find the reliable minimal passing distance earlier, thereby assisting in detecting potential risks timely. Fig. 11(c) further provides the ship pair collision risk evolution over time, where CR_{AB}* and CR_{AB} represent the collision risk that considers and not considers the influence of water topography. It is evident that CR_{AB}* provides more convincing assessment results since it can eliminate the estimated collision risk when the presence of islands obstructs the encountering ships. Consequently, the improved CPA-based method can detect collision risks reliably and timely with the consideration of both the potential ship dynamic motion and restricted water topography.

Secondly, the regional collision risk evaluation model is examined through two Axioms of sensitivity analysis [80–82]: (1) an

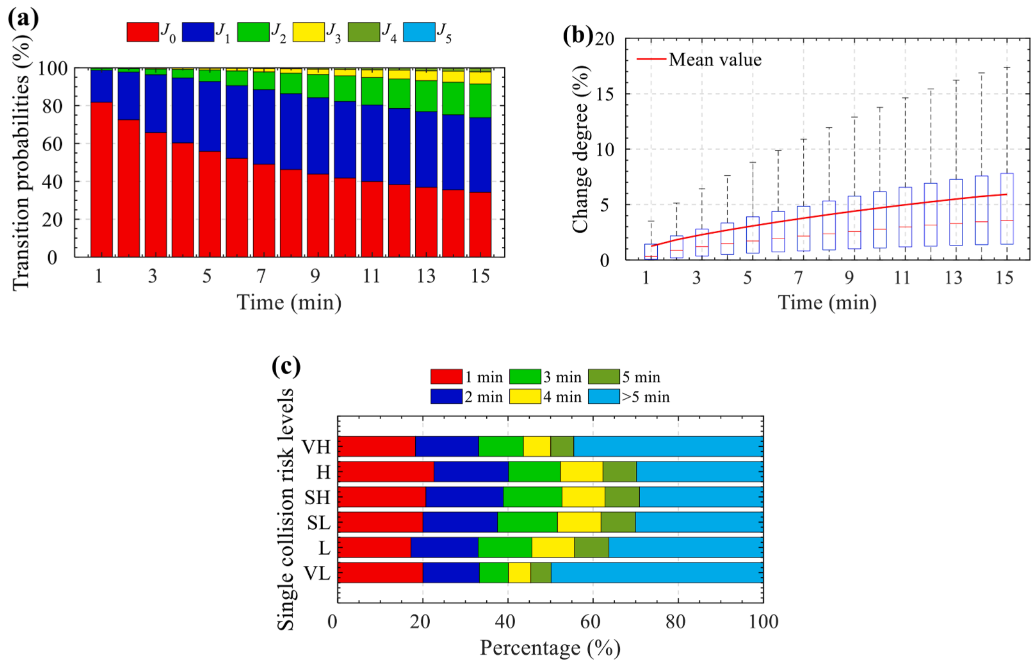

Fig. 9. Individual ship collision risk evolution. (a) Transition probabilities between different single ship collision risk levels over a short-term period; (b) change degree of SCR_i^2 over a short-term time; (c) life cycle of different single ship risk levels.

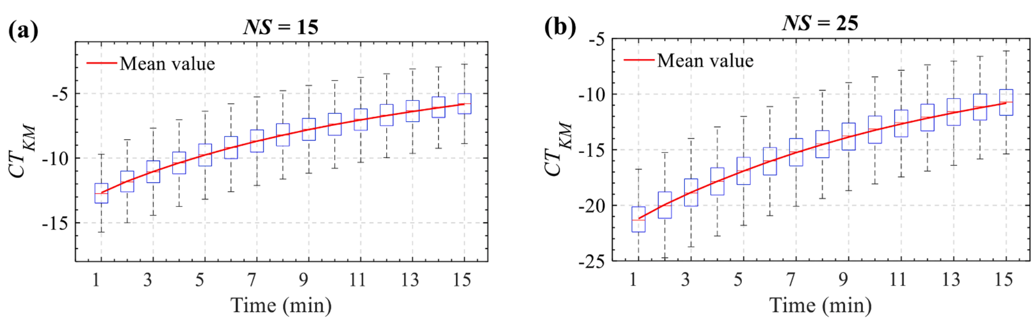

Fig. 10. Distance measure results between the current traffic partition and the future traffic partition with different time intervals. (a) $NS = 15$; (b) $NS = 25$.

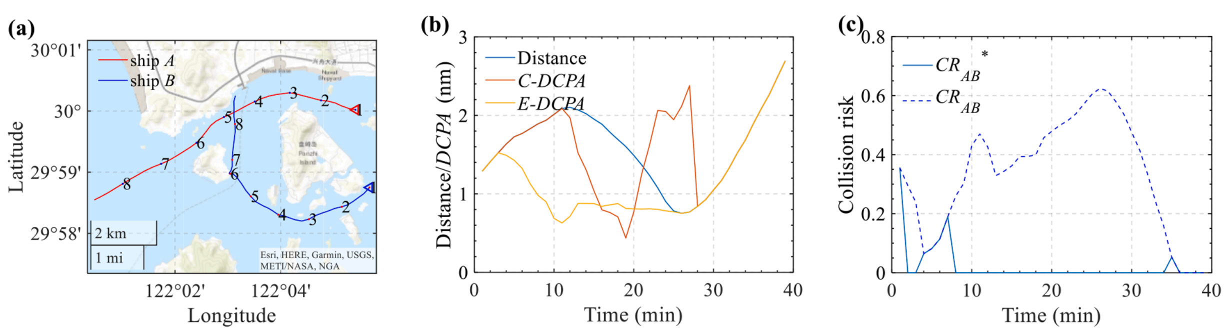

Fig. 11. A two-ship encounter scenario. (a) Ship trajectories; (2) evolutions of distance and DCPA indicators over time; (c) collision risk evolutions over time.

increase/decrease in the Number of Nodes (*NN*) or the Number of Edges (*NE*) in a traffic network scenario should lead to a corresponding increase/decrease in the regional collision risk *RCR*; (2) the total influence of *NN* and *NE* should not be smaller than the change by its subsets (i.e., any part of *NN* or *NE*). Following the two Axioms, Table 1 shows the effects of single factor change on the *RCR*. It is observed that the increase/decrease of *NN* or *NE* results in the correspondence change in the *RCR* and their change amplitude is positively correlated. These results are in good agreement with Axiom 1. Equally, Table 2 presents the effects of multiple factor change. There is a more substantial variation in the *RCR* when *NN* and *NE* change together. This coincides with Axiom 2, which proves the rationality of the regional collision risk model to some extent.

As for the validity examination of the individual collision risk model, the correlations between the *SCR* and the local network indices (i.e., D_i, VS_i, KS_i, CC_i) are investigated, as shown in Fig. 12. It is remarkable that the *SCR* has a significantly positive relation with each of local network indicators. These results conform to the common sense, i.e., a single ship

Table 1
Validity test (1) for regional collision risk model.

Change rate of *NN*	*ΔRCR*[1]	*ΔRCR*[2]	Change rate of *NE*	*ΔRCR*[1]	*ΔRCR*[2]
+5%	+0.22%	+0.74%	+5%	+2.2%	+2.95%
+10%	+0.38%	+1.49%	+10%	+4.57%	+5.93%
+15%	+0.55%	+2.22%	+15%	+6.86%	+8.97%
+20%	+0.75%	+2.96%	+20%	+9.12%	+12%
−5%	−5.99%	−8.07%	−5%	−4.93%	−6.55%
−10%	−12.84%	−16.25%	−10%	−9.92%	−12.58%
−15%	−21.14%	−24.77%	−15%	−15.08%	−18.28%
−20%	−28.24%	−32.83%	−20%	−20.14%	−23.58%

associated with larger local network indices should have a higher risk level than the one with smaller network indices. Simultaneously, the identical statistical analyses are carried out for the multi-ship collision risk model. According to Fig. 13, similar responses between the multi-ship collision risk (*MCR*) and the network indices can be observed, which helps further validate the model's feasibility and reliability.

Additionally, the robustness of the optimal maritime traffic partition model is tested. In fact, the performance of SNMF on graph partition has been comprehensively verified to be reliable and robust in [62,83,87]. Here, the proposed traffic partition model is compared with a widely favoured graph clustering algorithm (i.e., spectral clustering) to demonstrate its superiority and practical usability. Other clustering algorithms, such as density-based and prototype-based clustering, are not considered for comparison due to their limitations in coping with the complex interactions/interrelationships between any pair of data samples. Two graph-based metrics, 'NcutSilhouette' (*NcutS*) [47] and Normalized cut (*NC*) [64] are employed to check the traffic partition quality. Note that each produced cluster has a *NcutS* measure, and the overall partition performance is evaluated based on the average *NcutS* of all generated clusters in one traffic scenario. The comparable results are depicted in Fig. 14. It is noticeable that the developed model performs better than spectral clustering under various numbers of clusters, thereby demonstrating its generalization ability and stability. In addition, the statistical results show that the *NcutS* value of each generated traffic cluster is smaller than 1, indicating that all traffic scenarios are properly separated [47]. Therefore, the multiple dependent conflict relations among ships are well integrated when searching for the optimal scales for practical collision risk evaluation.

5. Conclusions

Multi-scale collision risk estimation regards different spatial scales as

Table 2
Validity test (2) for regional collision risk model.

Change rate of *NN*	+5%	+5%	−5%	−5%	+10%	+10%	−10%	−10%
Change rate of *NE*	/	+5%	/	−5%	/	+10%	/	−10%
ΔRCR[1]	+0.22%	+2.43%	−5.99%	−11%	+0.38%	+4.86%	−12.84%	−22.71%
ΔRCR[2]	+0.74%	+3.67%	−8.07%	−14.09%	+1.49%	+7.33%	−16.25%	−26.58%
Change rate of *NN*	+15%	+15%	−15%	−15%	+20%	+20%	−20%	−20%
Change rate of *NE*	/	+15%	/	−15%	/	+20%	/	−20%
ΔRCR[1]	+0.55%	+7.16%	−21.14%	−32.75%	+0.75%	+9.48%	−28.24%	−41.91%
ΔRCR[2]	+2.22%	+10.96%	−24.77%	−37.85%	+2.96%	+14.58%	−32.83%	−47.6%

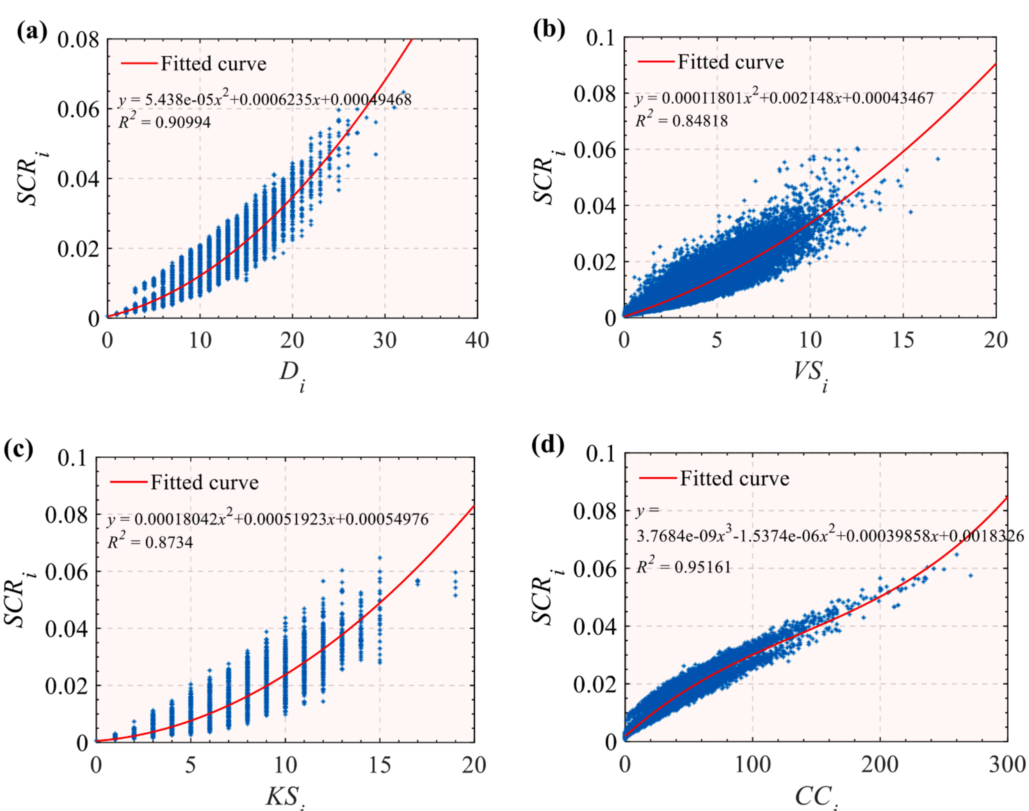

Fig. 12. Correlations between single ship collision risk and local network indices. (a) SCR_i vs. D_i; (b) SCR_i vs. VS_i; (c) SCR_i vs. KS_i; (d) SCR_i vs. CC_i.

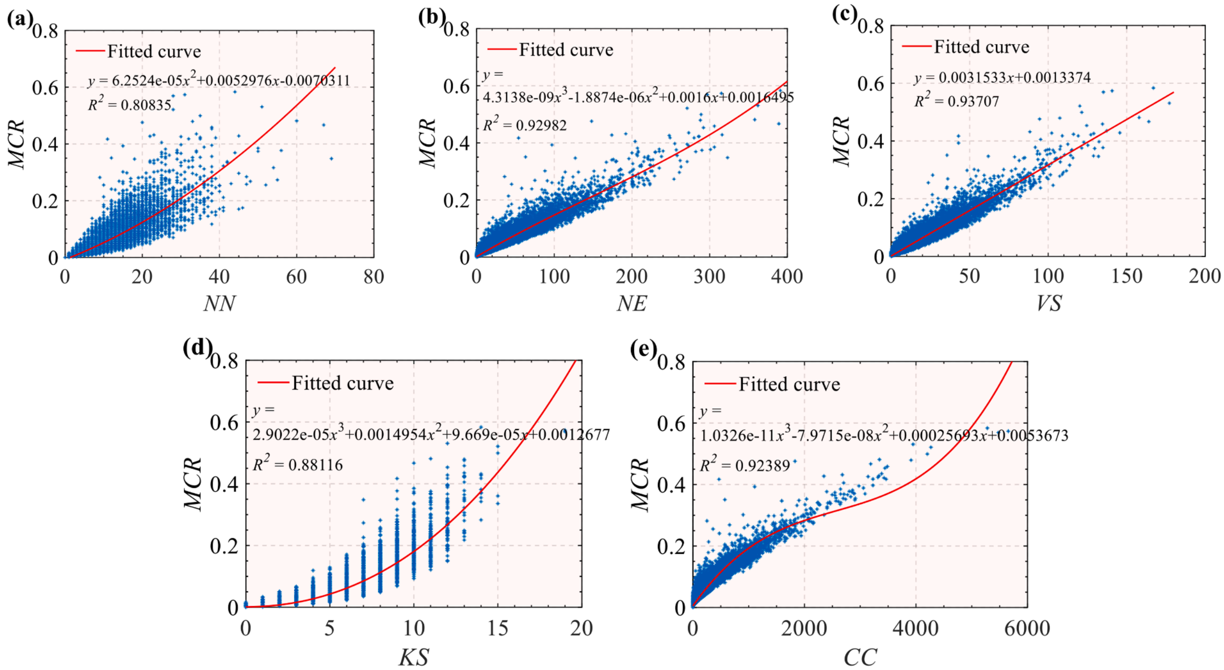

Fig. 13. Correlations between multi-ship collision risk and global network indices. (a) *MCR* vs. *NN*; (b) *MCR* vs. *NE*; (c) *MCR* vs. *VS*; (d) *MCR* vs. *KS*; (e) *MCR* vs. *CC*.

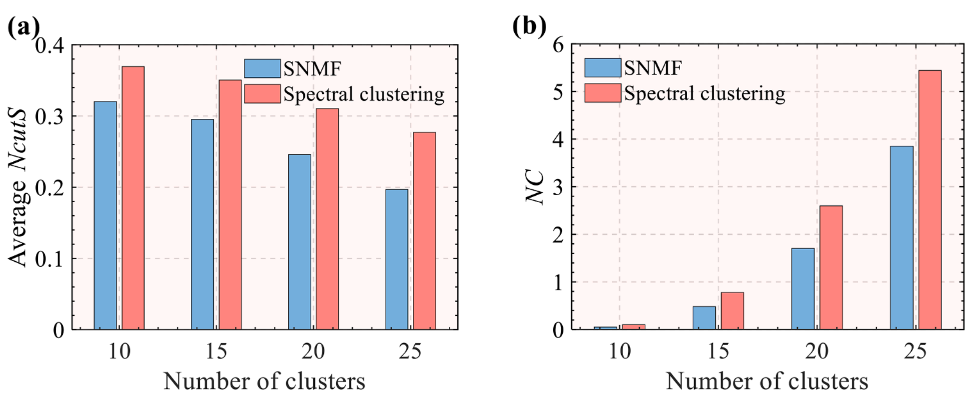

Fig. 14. Performance comparison between SNMF and spectral clustering: (a) average *NcutS* comparisons; (b) NC comparisons. A smaller average *NcutS* and NC indicate a better clustering performance.

Algorithm B1
Maritime traffic partition.

Input: The set of ships associated with their attributes $\{x_i\}_{i=1:N}$, and the desired number of clusters NS.
Output: The set of clusters $\{C_1, C_2, \cdots C_{NS}\}$.
// A. Similarity measure
1. Initialize similarity matrix as $W_{ij} \leftarrow 0_{[N\times N]}$.
2. For $\forall\, x_a, x_b \in \{x_i\}_{i=1:N}$ do
3. $W_{ab} = CR_{ab}$, where CR_{ab} is obtained by Eq. (1).
4. End
// B. SNMF implementation
5. $D = diag(d_i)$, where $d_i = \sum_{j=1}^{N} W_{ij}$.
6. $\widetilde{W} = D^{-1/2}WD^{-1/2}$
7. $H^* = \text{argmin} \parallel \widetilde{W} - HH^T \parallel^2$, where $H \in \mathbb{R}_+^{N\times NS}$.
8. For $i = 1 : N$ do
9. $j^* = \text{argmax}_{\{j=1,2,\cdots,NS\}} H_{ij}$
10. $x_i \in C_{j^*}$
11.End

different views to characterize the different aspects of a traffic scenario. This study shifts a paradigm of ship collision risk analysis from a single scale focused scheme towards a regium involving a multi-scale collision risk study to reveal the traffic risk patterns under different spatial granularity. It synergizes a series of techniques to achieve collision risk estimation at any spatial scale and capture the optimal spatial scope for risk analysis. The developed methodology has several unique features: (1) it incorporates the influence of ship motion dynamics and water topography on collision risk to ensure the applicability in generalized scenarios; (2) the topological characteristics of multiple ship conflicts are explicitly considered to reveal the resolving difficulty of collisions brought by traffic interaction structure; (3) it pioneers the application of node deletion method to quantify the aggregation risk criticality of any multiple ship interactions to a regional traffic situation; (4) a competitive SNMF framework is embedded to search for the optimal traffic clusters at any spatial scale. Comprehensive experiments based on real-AIS data are performed to evaluate and check the performance of the proposed approach. Experimental results show that the proposed methodology can offer a complete comprehension for a traffic scenario and facilitate strategic maritime safety management. Additionally, the robustness and superiority of the proposed methods are tested and examined through sensitivity analysis, correlation analysis and model comparison. This study therefore could be used in practise to support intelligent maritime perception and promote maritime system automation.

Future research will put efforts on the following aspects. First, the impact of additional factors, e.g., ship maneuverability, human behaviours, and environmental disturbances, on the traffic collision risk could be taken into consideration. It could help improve the collision risk evaluation accuracy. Second, the propagation and prediction of maritime traffic risk deserve further concern. It will enhance the perception ability for the forthcoming traffic situations, which is helpful for issuing an early collision alert and preventing the time lag in risk management response. Finally, a new conflict resolution approach that can coordinate the local and regional collision risk could be beneficial to guide surveillance operators to devise multi-layered strategies for hierarchical risk control purposes.

CRediT authorship contribution statement

Xuri Xin: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Kezhong Liu:** Supervision, Resources, Project administration, Funding acquisition, Data curation, Conceptualization. **Sean Loughney:** Writing – review & editing, Supervision, Project administration. **Jin Wang:** Writing – review & editing, Supervision, Project administration. **Huanhuan Li:** . **Nduka Ekere:** Writing – review & editing, Project administration. **Zaili Yang:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

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.

Data availability

The authors do not have permission to share data.

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant No. 52031009) and a European Research Council project under the European Union's Horizon 2020 research and innovation programme (TRUST CoG 2019 864724).

APPENDIX A. Detailed illustration of the FCI approach

The detailed procedure for implementing the FCI model is depicted as follows.

Suppose there are n assessment samples and each sample records m indices as $A_j = (a_{1j}, a_{2j}, \ldots, a_{mj})$. Then the sample dataset is expressed as matrix $A = (a_{ij})_{m \times n}$, where a_{ij} represents the ith index of sample j.

As each index has different orders of magnitude, the elements in A should be standardized to eliminate the dimensionality influence, using the following equation:

$$r_{ij} = \left(a_{ij} - a_{i,min}\right) / \left(a_{i,max} - a_{i,min}\right) \tag{A.1}$$

where $a_{i,min}$ and $a_{i,max}$ represent the minimum and maximum values in the ith row in A, respectively. Hence, matrix A can be transformed into a normalised matrix R.

After that, assume that the n samples with m attributes are clustered with c classes/patterns, the membership matrix and class center matrix can be defined as $U = (u_{kj})_{c \times n}$ and $S = (s_{ik})_{m \times c}$, where u_{kj} represents the membership value of sample j assigned to class k, subject to $0 \leq u_{kj} \leq 1$ and $\sum_{k=1}^{c} u_{kj} = 1$, and s_{ik} denotes the center of index i in class k, satisfying $0 \leq s_{ik} \leq 1$.

To gain the optimal U and S, the objective function that minimizes the square sum of the weighted Euclidean distance from samples 1-n to class centres 1-c is constructed as follows:

$$\min\left[F\left(w_i, u_{kj}, s_{ik}\right)\right] = \min\left\{\sum_{j=1}^{n}\sum_{k=1}^{c}\left(u_{kj}^2\sum_{i=1}^{m}\left(w_i\left(r_{ij} - s_{ik}\right)\right)^2\right)\right\} \tag{A.2}$$

where w_i represents the weight of different indices, subject to $0 \leq w_i \leq 1$ and $\sum_{i=1}^{m} w_i = 1$.

According to the objective function in Eq. (A.2), the w, U, and S can be iteratively optimized by the Lagrange multiplier method, using the following equations:

$$w_i = \left[\sum_{i=1}^{m}\frac{\sum_{j=1}^{n}\sum_{k=1}^{c}\left[u_{kj}\left(r_{ij} - s_{ik}\right)\right]^2}{\sum_{j=1}^{n}\sum_{k=1}^{c}\left[u_{kj}\left(r_{ij} - s_{ik}\right)\right]^2}\right]^{-1} \tag{A.3}$$

$$u_{kj} = \left[\sum_{h=1}^{c}\frac{\sum_{i=1}^{m}\left[w_i\left(r_{ij} - s_{ik}\right)\right]^2}{\sum_{i=1}^{m}\left[w_i\left(r_{ij} - s_{ih}\right)\right]^2}\right]^{-1} \tag{A.4}$$

$$s_{ik} = \sum_{j=1}^{n} u_{kj}^2 w_i^2 r_{ij} \Big/ \sum_{j=1}^{n} u_{kj}^2 w_i^2 \tag{A.5}$$

The specific update process comprises the following steps.

(1) Initialize the precision parameters ε_1, ε_2, and ε_3 used for w_i, u_{kj}, and s_{ik}.
(2) Let $l = 0$ and generate the original w^l and U^l which satisfy the constraints mentioned above.

(3) Calculate the original S^l by inputting the original w^l and U^l into Eq. (A.5) and $l = l + 1$.
(4) Update w^l, U^l, and S^l via Eqs. (A.3)–(A.5), respectively.
(5) Identify whether all the following constraints are satisfied

$$\max_i \left| w_i^{l+1} - w_i^l \right| \le \varepsilon_1$$

$$\max_{kj} \left| u_{kj}^{l+1} - u_{kj}^l \right| \le \varepsilon_2$$

$$\max_{ik} \left| s_{ik}^{l+1} - s_{ik}^l \right| \le \varepsilon_3$$

If not, $l = l + 1$ and repeat step 4 until the above termination conditions are held.

(6) Output the optimal w, U, and S.

APPENDIX B. Pseudocode for the traffic partition procedure

Algorithm B.1 presents the pseudocode of the maritime traffic partition. It involves two important modules: similarity measure and SNMF implementation. The similarity measure takes the collision risk as the similarity of ship pairs (step 3). Further, the Newton-like algorithm is used to optimize the SNMF framework in Eq. (10) (step 7). This algorithm can achieve higher accuracy and quality when solving small size optimization problems (e.g., the number of samples is smaller than 3000) [62]. Besides, it is implemented many times with different initial H to avoid falling into local minima. The pseudocode of the Newton-like algorithm can be found in [48,83].

APPENDIX C. Training results of the FCI model

This appendix lists the training results of the FCI model. A total of 4315 traffic scenarios are used as the training samples, in which each of them is extracted every 10 min from one month of AIS data. One well-known problem that needs to be addressed is identifying the optimal number of levels/classes of regional collision risk. Here, six validity indices in [84], including partition coefficient (PC), partition entropy (PE), modified partition coefficient (MPC), VFS, VXB, and VK, are adopted to measure the performance of the FCI model. In these indices, a high PC and MPC value indicates that the dataset is well clustered. In contrast, a small value of the remaining indices means that a good traffic partition is produced. Fig. C.1 illustrates the average validity performance of the six indices, in which each of the number of risk levels is run 20 times with randomly sampled initial w and U. It is seen that the MPC, VXB, and VK indices have the best performance when the number of risk levels is six. Consequently, the optimal number of regional risk levels is set to be six. More detailed explanations about the adopted validity indices are summarized in [84].

Fig. C2 further displays the optimal results of the FCI model when the number of risk levels is six. By inputting the optimal w and S into Eq. (A.4), the regional risk membership distribution of the new traffic scenarios can be calibrated.

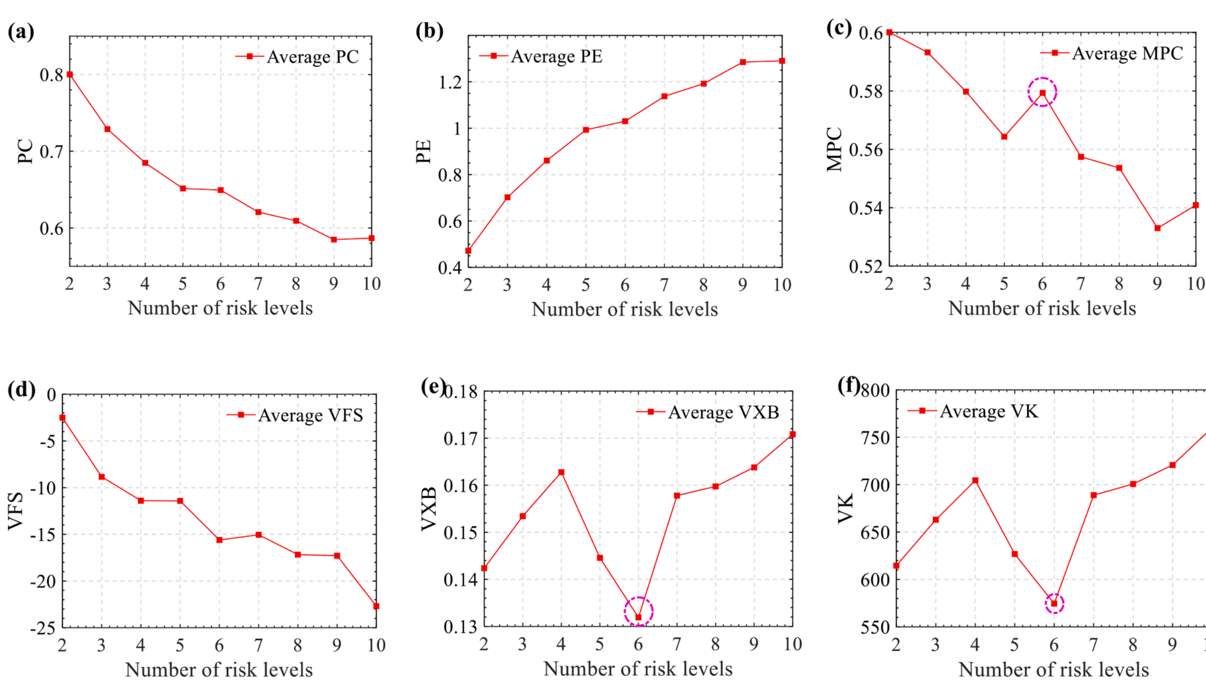

Fig. C1. Model performance illustration with different numbers of risk levels.

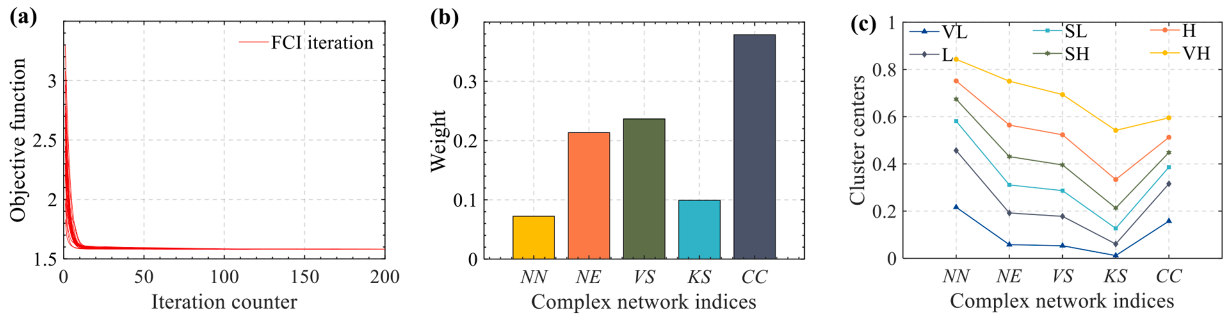

Fig. C2. Optimal results of FCI model. (a) Iteration optimization processes of FCI with 20 randomly initial *w* and *S*; (b) weights of five adopted network indices; (c) clustering centres of six regional collision risk levels.

References

[1] Yu H, Murray AT, Fang Z, Liu J, Peng G, Solgi M, et al. Ship path optimization that accounts for geographical traffic characteristics to increase maritime port safety. IEEE Trans Intell Transp Syst 2021;23:5765–76.
[2] Liu RW, Liang M, Nie J, Lim WYB, Zhang Y, Guizani M. Deep learning-powered vessel trajectory prediction for improving smart traffic services in maritime internet of things. IEEE Trans Netw Sci Eng 2022;9(5):3080–94.
[3] Liang M, Liu RW, Zhan Y, Li H, Zhu F, Wang FY. Fine-grained vessel traffic flow prediction with a Spatio-temporal multigraph convolutional network. IEEE Trans Intell Transp Syst 2022;23(12):23694–707.
[4] Register L. Cyber-enabled ships shipright procedure assignment for cyber descriptive notes for autonomous & remote access ships. Lloyd's regist guid doc version 20 2017.
[5] Rivkin BS. Unmanned ships: navigation and more. Gyroscopy Navig 2021;12:96–108.
[6] Bakdi A, Glad IK, Vanem E. Testbed scenario design exploiting traffic big data for autonomous ship trials under multiple conflicts with collision/grounding risks and Spatio-temporal dependencies. IEEE Trans Intell Transp Syst 2021;22:7914–30.
[7] Ashraf I, Park Y, Hur S, Kim SW, Alroobaea R, Bin ZY, et al. A survey on cyber security threats in IoT-enabled maritime industry. IEEE Trans Intell Transp Syst 2022;24(2):2677–90.
[8] Zhang M, Kujala P, Hirdaris S. A machine learning method for the evaluation of ship grounding risk in real operational conditions. Reliab Eng Syst Saf 2022;226:108697.
[9] Zhang M, Montewka J, Manderbacka T, Kujala P, Hirdaris S. A big data analytics method for the evaluation of ship-ship collision risk reflecting hydrometeorological conditions. Reliab Eng Syst Saf 2021;213:107674.
[10] Tian YF, Chen LJ, Huang LW, Mou JM. Featured risk evaluation of nautical navigational environment using a risk cloud model. J Mar Eng Technol 2020;19:115–29.
[11] Fang Z, Yu H, Ke R, Shaw SL, Peng G. Automatic identification system-based approach for assessing the near-miss collision risk dynamics of ships in ports. IEEE Trans Intell Transp Syst 2018;20:534–43.
[12] Yu H, Fang Z, Murray AT, Peng G. A direction-constrained space-time prism-based approach for quantifying possible multi-ship collision risks. IEEE Trans Intell Transp Syst 2019;22:131–41.
[13] Thombre S, Zhao Z, Ramm-Schmidt H, García JMV, Malkamäki T, Nikolskiy S, et al. Sensors and AI techniques for situational awareness in autonomous ships: a review. IEEE Trans Intell Transp Syst 2020;23(1):64–83.
[14] Chen P, Huang Y, Mou J, van Gelder P. Probabilistic risk analysis for ship-ship collision: state-of-the-art. Saf Sci 2019;117:108–22.
[15] Huang Y, Chen L, Chen P, Negenborn RR, van Gelder P. Ship collision avoidance methods: state-of-the-art. Saf Sci 2020;121:451–73.
[16] Yang D, Wu L, Wang S, Jia H, Li KX. How big data enriches maritime research–a critical review of automatic identification system (AIS) data applications. Transp Rev 2019;39:755–73.
[17] Tu E, Zhang G, Rachmawati L, Rajabally E, Huang GB. Exploiting AIS data for intelligent maritime navigation: a comprehensive survey from data to methodology. IEEE Trans Intell Transp Syst 2017;19:1559–82.
[18] Xin X, Liu K, Yang Z, Zhang J, Wu X. A probabilistic risk approach for the collision detection of multi-ships under spatiotemporal movement uncertainty. Reliab Eng Syst Saf 2021;215:107772.
[19] Huang Y, van Gelder P. Time-varying risk measurement for ship collision prevention. Risk Anal 2020;40:24–42.
[20] Li M, Mou J, Chen P, Rong H, Chen L, van Gelder P. Towards real-time ship collision risk analysis: an improved R-TCR model considering target ship motion uncertainty. Reliab Eng Syst Saf 2022;226:108650.
[21] Chen P, Huang Y, Mou J, van Gelder P. Ship collision candidate detection method: a velocity obstacle approach. Ocean Eng 2018;170:186–98.
[22] Zhang W, Goerlandt F, Montewka J, Kujala P. A method for detecting possible near miss ship collisions from AIS data. Ocean Eng 2015;107:60–9.
[23] Szlapczynski R, Szlapczynska J. Review of ship safety domains: models and applications. Ocean Eng 2017;145C:277–89.
[24] Zhao Y, Li W, Shi P. A real-time collision avoidance learning system for unmanned surface vessels. Neurocomputing 2016;182:255–66.
[25] Gil M, Montewka J, Krata P, Hinz T, Hirdaris S. Determination of the dynamic critical maneuvering area in an encounter between two vessels: operation with negligible environmental disruption. Ocean Eng 2020;213:107709.
[26] Gil M, Kozioł P, Wróbel K, Montewka J. Know your safety indicator–a determination of merchant vessels bow crossing range based on big data analytics. Reliab Eng Syst Saf 2022;220:108311.
[27] Gil M. A concept of critical safety area applicable for an obstacle-avoidance process for manned and autonomous ships. Reliab Eng Syst Saf 2021;214:107806.
[28] Montewka J, Hinz T, Kujala P, Matusiak J. Probability modeling of vessel collisions. Reliab Eng Syst Saf 2010;95:573–89.
[29] Xiao Z, Fu X, Zhang L, Zhang W, Liu RW, Liu Z, et al. Big data driven vessel trajectory and navigating state prediction with adaptive learning, motion modeling and particle filtering techniques. IEEE Trans Intell Transp Syst 2020;23:3696–709.
[30] Zhang W, Feng X, Qi Y, Feng S, Zhang Y, Wang Y. Towards a model of regional vessel near-miss collision risk assessment for open waters based on AIS data. J Navig 2019;72:1449–68.
[31] Liu Z, Wu Z, Zheng Z. A cooperative game approach for assessing the collision risk in multi-vessel encountering. Ocean Eng 2019;187:106175.
[32] Zhang M, Zhang D, Fu S, Kujala P, Hirdaris SA. Predictive analytics method for maritime traffic flow complexity estimation in inland waterways. Reliab Eng Syst Saf 2022;220:108317.
[33] Xin X, Yang Z, Liu K, Zhang J, Wu X. Multi-stage and multi-topology analysis of ship traffic complexity for probabilistic collision detection. Expert Syst Appl 2022;213:118890.
[34] Bombelli A, Santos BF, Tavasszy L. Analysis of the air cargo transport network using a complex network theory perspective. Transp Res Part E Logist Transp Rev 2020;138:101959.
[35] Radanovic M, Eroles MAP, Koca T, Gonzalez JJR. Surrounding traffic complexity analysis for efficient and stable conflict resolution. Transp Res Part C Emerg Technol 2018;95:105–24.
[36] Wang H, Song Z, Wen R, Zhao Y. Study on evolution characteristics of air traffic situation complexity based on complex network theory. Aerosp Sci Technol 2016;58:518–28.
[37] Zhou Y, Kundu T, Qin W, Goh M, Sheu JB. Vulnerability of the worldwide air transportation network to global catastrophes such as COVID-19. Transp Res Part E Logist Transp Rev 2021;154:102469.
[38] Wen X, Tu C, Wu M. Node importance evaluation in aviation network based on "No Return" node deletion method. Phys A Stat Mech Appl 2018;503:546–59.
[39] Liu H, Tian Z, Huang A, Yang Z. Analysis of vulnerabilities in maritime supply chains. Reliab Eng Syst Saf 2018;169:475–84.
[40] Liu K, Yuan Z, Xin X, Zhang J, Wang W. Conflict detection method based on dynamic ship domain model for visualization of collision risk hot-spots. Ocean Eng 2021;242:110143.
[41] Rong H, Teixeira AP, Soares CG. Spatial correlation analysis of near ship collision hotspots with local maritime traffic characteristics. Reliab Eng Syst Saf 2021;209:107463.
[42] Yoo SL. Near-miss density map for safe navigation of ships. Ocean Eng 2018;163:15–21.
[43] Xin X, Liu K, Loughney S, Wang J, Yang Z. Maritime traffic clustering to capture high-risk multi-ship encounters in complex waters. Reliab Eng Syst Saf 2022;230:108936.
[44] Zhen R, Shi Z, Shao Z, Liu J. A novel regional collision risk assessment method considering aggregation density under multi-ship encounter situations. J Navig 2021;75(1):76–94.
[45] Saeedmanesh M, Geroliminis N. Clustering of heterogeneous networks with directional flows based on "Snake" similarities. Transp Res Part B Methodol 2016;91:250–69.
[46] Saeedmanesh M, Geroliminis N. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. Transp Res Procedia 2017;23:962–79.
[47] Ji Y, Geroliminis N. On the spatial partitioning of urban transportation networks. Transp Res Part B Methodol 2012;46:1639–56.
[48] Gu Z, Saberi M. A bi-partitioning approach to congestion pattern recognition in a congested monocentric city. Transp Res Part C Emerg Technol 2019;109:305–20.

[49] Zhang J, Zhang D, Yan X, Haugen S, Soares CG. A distributed anti-collision decision support formulation in multi-ship encounter situations under COLREGs. Ocean Eng 2015;105:336–48.
[50] Xiao Z, Fu X, Zhang L, Goh RSM. Traffic pattern mining and forecasting technologies in maritime traffic service networks: a comprehensive survey. IEEE Trans Intell Transp Syst 2019;21:1796–825.
[51] Osekowska E, Johnson H, Carlsson B. Grid size optimization for potential field based maritime anomaly detection. Transp Res Procedia 2014;3:720–9.
[52] Kim KI, Jeong JS, Park GK. Development of a gridded maritime traffic DB for e-navigation. Int. J. eNavig. Marit. Econ. 2014;1:39–47.
[53] Wang H, Song Z, Wen R. Modeling air traffic situation complexity with a dynamic weighted network approach. J Adv Transp 2018;2018:1–15.
[54] Wu X, Liu K, Zhang J, Yuan Z, Liu J, Yu Q. An optimized collision avoidance decision-making system for autonomous ships under human-machine cooperation situations. J Adv Transp 2021;2021:1–17.
[55] Liu J, Zhang J, Yan X, Soares CG. Multi-ship collision avoidance decision-making and coordination mechanism in mixed navigation scenarios. Ocean Eng 2022;257:111666.
[56] Zekun W, Xiangxi W, Minggong W. Identification of key nodes in aircraft state network based on complex network theory. IEEE Access 2019;7:60957–67.
[57] Lyu HM, Shen SL, Zhou A, Zhou WH. Data in flood risk assessment of metro systems in a subsiding environment using the interval FAHP–FCA approach. Data Br 2019;26:104468.
[58] He Y, Zhou J, Kou P, Lu N, Zou Q. A fuzzy clustering iterative model using chaotic differential evolution algorithm for evaluating flood disaster. Expert Syst Appl 2011;38:10060–5.
[59] He Y, Wan J. Flood hazard assessment based on fuzzy clustering iterative model and chaotic particle swarm optimization. J Ambient Intell Humaniz Comput 2021;12(1):933–42.
[60] Wang Q, Yang X. Investigating the sustainability of renewable energy–an empirical analysis of European Union countries using a hybrid of projection pursuit fuzzy clustering model and accelerated genetic algorithm based on real coding. J Clean Prod 2020;268:121940.
[61] Wang J, Wu X, Yan B, Guo J. Improved method of node importance evaluation based on node contraction in complex networks. Procedia Eng 2011;15:1600–4.
[62] Kuang D, Yun S, Park H. SymNMF: nonnegative low-rank approximation of a similarity matrix for graph clustering. J Glob Optim 2015;62:545–74.
[63] Ding C, He X, Simon HD. On the equivalence of nonnegative matrix factorization and spectral clustering. In: Proceedings of the SIAM international conference on data mining, SIAM; 2005. p. 606–10.
[64] Shi J, Malik J. Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 2000;22:888–905.
[65] Rong H, Teixeira AP, Soares CG. Maritime traffic probabilistic prediction based on ship motion pattern extraction. Reliab Eng Syst Saf 2022;217:108061.
[66] Rong H, Teixeira AP, Soares CG. Ship trajectory uncertainty prediction based on a Gaussian process model. Ocean Eng 2019;182:499–511.
[67] Fuentes G. Generating bunkering statistics from AIS data: a machine learning approach. Transp Res Part E Logist Transp Rev 2021;155:102495.
[68] Xin X, Liu K, Yang X, Yuan Z, Zhang J. A simulation model for ship navigation in the "Xiazhimen" waterway based on statistical analysis of AIS data. Ocean Eng 2019;180:279–89.
[69] Yu Q, Liu K, Teixeira AP, Soares CG. Assessment of the influence of offshore wind farms on ship traffic flow based on AIS data. J Navig 2020;73:131–48.
[70] Zhang J, Liu J, Hirdaris S, Zhang M, Tian W. An interpretable knowledge-based decision support method for ship collision avoidance using AIS data. Reliab Eng Syst Saf 2023;230:108919.
[71] Zhang M, Taimuri G, Zhang J, Hirdaris S. A deep learning method for the prediction of 6-DoF ship motions in real conditions. Proc Inst Mech Eng Part M J Eng Marit Environ 2023:14750902231157852.
[72] Li H, Lam JSL, Yang Z, Liu J, Liu RW, Liang M, et al. Unsupervised hierarchical methodology of maritime traffic pattern extraction for knowledge discovery. Transp Res Part C Emerg Technol 2022;143:103856.
[73] Li H, Liu J, Yang Z, Liu RW, Wu K, Wan Y. Adaptively constrained dynamic time warping for time series classification and clustering. Inf Sci 2020;534:97–116 (Ny).
[74] Liang M, Liu RW, Li S, Xiao Z, Liu X, Lu F. An unsupervised learning method with convolutional auto-encoder for vessel trajectory similarity computation. Ocean Eng 2021;225:108803.
[75] Kang L, Meng Q, Liu Q. Fundamental diagram of ship traffic in the Singapore Strait. Ocean Eng 2018;147:340–54.
[76] Zhao L, Shi G, Yang J. Ship trajectories pre-processing based on AIS data. J Navig 2018;71:1210–30.
[77] Zhang L, Meng Q, Fwa TF. Big AIS data based spatial-temporal analyses of ship traffic in Singapore port waters. Transp Res Part E Logist Transp Rev 2019;129:287–304.
[78] Chi Y, Song X, Zhou D, Hino K, Tseng BL. Evolutionary spectral clustering by incorporating temporal smoothness. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining; 2007. p. 153–62.
[79] Du L, Goerlandt F, Banda OAV, Huang Y, Wen Y, Kujala P. Improving stand-on ship's situational awareness by estimating the intention of the give-way ship. Ocean Eng 2020;201:107110.
[80] Yu Q, Liu K, Chang CH, Yang Z. Realising advanced risk assessment of vessel traffic flows near offshore wind farms. Reliab Eng Syst Saf 2020;203:107086.
[81] Fan S, Blanco-Davis E, Yang Z, Zhang J, Yan X. Incorporation of human factors into maritime accident analysis using a data-driven Bayesian network. Reliab Eng Syst Saf 2020;203:107070.
[82] Zhang D, Yan XP, Yang ZL, Wall A, Wang J. Incorporation of formal safety assessment and Bayesian network in navigational risk estimation of the Yangtze River. Reliab Eng Syst Saf 2013;118:93–105. https://doi.org/10.1016/j.ress.2013.04.006.
[83] Kuang D, Ding C, Park H. Symmetric nonnegative matrix factorization for graph clustering. In: Proceedings of the SIAM international conference on data mining, SIAM; 2012. p. 106–17.
[84] Wang W, Zhang Y. On fuzzy cluster validity indices. Fuzzy Sets Syst 2007;158:2095–117.
[85] Li H, Yang Z. Incorporation of AIS data-based machine learning into unsupervised route planning for maritime autonomous surface ships. Transp Res Part E Logist Transp Rev 2023;176:103171.
[86] Li H, Jiao H, Yang Z. AIS data-driven ship trajectory prediction modelling and analysis based on machine learning and deep learning methods. Transp Res Part E Logist Transp Rev 2023;175:103152.
[87] Xin X, Liu K, Loughney S, Wang J, Li H, Yang Z. Graph-based ship traffic partitioning for intelligent maritime surveillance in complex port waters. Expert Syst Appl 2023:120825.