Ma, Quandang; Tang, Huan; Liu, Cong; Zhang, Mingyang; Zhang, Dingze; Liu, Zhao; Zhang, Liye

A big data analytics method for the evaluation of maritime traffic safety using automatic identification system data

Published in:
OCEAN AND COASTAL MANAGEMENT

DOI:
10.1016/j.ocecoaman.2024.107077

Published: 01/05/2024

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
A big data analytics method for the evaluation of maritime traffic safety using automatic identification system data

Quandang Ma a,b, Huan Tang a, Cong Liu c,*, Mingyang Zhang c, Dingze Zhang a, Zhao Liu a, Liye Zhang d

a School of Navigation, Wuhan University of Technology, Wuhan, China
b National Engineering Research Center for Water Transport Safety, Wuhan University of Technology, Wuhan, China
c Department of Mechanical Engineering, Marine Technology Group, Aalto University, Espoo, Finland
d College of Transportation, Shandong University of Science and Technology, Qingdao, China

ARTICLE INFO

Keywords:
Big-data analytics
Maritime traffic safety
Spatial-temporal analysis
Yangtze river
Maritime accidents

ABSTRACT

The complex traffic situations are among the factors influencing maritime safety. They can be quantitatively estimated through the analysis of traffic data. This paper explores the impact of complex traffic situations on maritime safety, focusing on inland waterway traffic. It presents a big data analytics method, utilizing data from the Automatic Identification System (AIS) and historical maritime accident records. The methodology involves AIS data preprocessing and spatial autocorrelation models, including Moran’s index, to extract and evaluate the dynamic characteristics of maritime traffic. The analysis of traffic characteristic includes a thorough investigation into the spatial-temporal distribution of ship average speed and trajectory density. The paper then introduces an effective traffic characteristic analysis model that evaluates the relationship between maritime traffic patterns and accidents. The study, specifically targeting the Nanjing section of the Yangtze River, reveals variations in ship trajectory density and average speed over time. It identifies several hotspots with a significant local correlation between these factors. Moreover, a substantial correlation is found between the locations of maritime accidents and areas with increased ship trajectory density and average speed. These results may provide insights for traffic safety management and highlight strategies for preventing maritime accidents.

1. Introduction

The burgeoning global economy and its expanding international influence have catalyzed significant advancements in maritime transportation, affirming its status as a pivotal transportation mode (e.g., Deng et al., 2021; Xu et al., 2022; Shi et al., 2023; Li et al., 2023; Yi et al., 2023). However, such growth has led to greater diversity and an increased number of ships navigating inland waterways (Chai et al., 2020; Kum and Sahin, 2015; Lin et al., 2021; Wang and Cullinane, 2014), resulting in more complex traffic scenarios and greater navigational risks (e.g., Bye and Aalberg, 2018; Zhang et al., 2013; Fu et al., 2022a,b). Consequently, maritime accidents have seen a significant rise (e.g., Christian and Kang, 2017; Mou et al., 2019; Liu et al., 2023; Fu et al., 2022a,b). An examination of accident records underscores a notable frequency of maritime accidents within the Yangtze River, particularly in the Nanjing and Nantong sections. As shown in Fig. 1, each of these two regions experienced over 300 accidents. It’s imperative to develop a robust methodology for analyzing and improving traffic safety in inland waterways. The urgent need to enhance traffic safety in inland waterways has led to scholarly efforts focused on mitigating maritime traffic risks (e.g., Szlapczynski and Szlapczynska, 2017; Rong et al., 2022).

1.1. Literature review

The criticality of analyzing maritime traffic for the purpose of enhancing maritime traffic safety evaluation is well-documented in contemporary research (e.g., Fan et al., 2020a,b; Luo and Shin, 2019). Existing scholarly endeavors in this domain predominantly revolve around two primary modeling approaches: (a) semi-empirical models, and (b) data-driven models. This section offers a comprehensive literature review of both methodologies. Semi-empirical models typically combine empirical observations with theoretical constructs, providing a balanced perspective that is rooted in practical maritime scenarios while...
being guided by theoretical principles. Data-driven models leverage the burgeoning volume of maritime data, harnessing advanced computational techniques to extrapolate patterns and insights that are not immediately apparent through traditional analysis.

1.1.1. Semi-empirical models

The semi-empirical models help estimate the probability and consequence of accidents on the basis of accident data statistics and expert judgment. They can be divided into 8 categories based on the modelling techniques, as shown in Table 1. Common modelling tools include Bayesian Networks models (Fu et al., 2023), Fault Tree and Event Tree models (Zhang et al., 2019), Failure Mode Effect Analysis (Başhan et al., 2020), Fuzzy Logic methods (Wang et al., 2014), System-Theoretic Accident Model and Processes (Pü, et al., 2022a,b), Maritime Traffic Simulation tools (Weng et al., 2020), Waterway Geometrical models (Zhang et al., 2022a,b), and others (e.g., Luo and Shin, 2019; Fujii and Tanaka, 1971).

Overall, semi-empirical models excel at assessing maritime risks and identifying high-risk areas or factors in specific scenarios from a macro perspective. This offers valuable theoretical insights for traffic managers and policymakers. However, these models’ significant limitation lies in their neglect of dynamic elements, such as variable traffic patterns over time. This results in an inability to identify emergent risk scenarios and provide effective risk mitigation strategies for managing traffic safety in dynamically changing traffic conditions.

1.1.2. Data-driven models

To address the inherent subjectivity and data gaps in semi-empirical models, many researchers are now focusing on data-driven models for maritime risk analysis. Liu et al. (2021) and others have demonstrated that these models are effective tools for maritime accident prevention and safety management. As shown in Table 2, data-driven models necessitate extensive historical data, including AIS data, maritime accident records, ship-specific information, and others.

Technological advancements now facilitate the direct acquisition of ship data through Vessel Traffic Services (VTS) and AIS, yielding more comprehensive and reliable datasets (Wolsing et al., 2022). This supports the notion that AIS data-based studies offer a more scientific and forward-thinking approach to ship traffic analysis (Shelmerdine, 2015). AIS data-driven risk evaluation models can be broadly categorized into static and dynamic traffic evaluation methods (e.g., Yan et al., 2022; Zhang et al., 2022a,b; Fagerholt et al., 2015; Huang et al., 2023). Static methods primarily focus on the geographical location, type, and size of ships, employing techniques like geographical visualization of AIS data (He et al., 2021) and ArcGIS-based trajectory displays (Tsou, 2010). However, these methods often overlook the interplay of ship characteristics with spatial-temporal dimensions, thus limiting their efficacy in managing ship traffic safety.

The nature of maritime traffic demands advanced approaches for evaluation, which has spurred extensive research into dynamic traffic evaluation methods. These methods focus on analyzing key elements such as ship average speed, heading, and position, integrating both temporal and spatial data for a more nuanced analysis. This approach acknowledges that ship traffic conditions are not only contingent on its geographic location but also subject to continuous and unpredictable changes over time, a concept supported by studies from Zhang et al. (2022a,b), Zhang et al. (2019), and Yang et al. (2013). Pioneering research by Jiao et al. (2016), Wang et al. (2014), and Başhan et al. (2020) has furthered our understanding of maritime transportation risk by considering the dynamic aspects of maritime traffic. A notable contribution in this field is by Im and Luong (2019), who employed a potential risk ship domain model to evaluate how ship length, speed, and navigational conditions affect the ship domain’s size and shape. Their model also enabled the determination of the ship domain size and the associated potential collision risk index at various risk levels. Additionally, Balmat et al. (2011), Chai et al. (2017), and Oviedo-Trespalacios et al. (2017) conducted in-depth research into the interplay among the number of maritime accidents, their severity, the resultant losses, and their specific locations. These studies elucidated how these factors collectively impact maritime transportation safety.

Comprehensively reviewing maritime risk assessment literature reveals a diverse array of content and methodologies, each with its
inherent strengths and limitations. Semi-empirical models, often hinges on the subjective judgment of decision-makers, typically engage in macro-level and qualitative analyses. This approach, while broad in scope, can compromise the precision of risk assessments and may fall short in providing nuanced guidance for decision-makers (e.g., Fan et al., 2020a,b; Xue et al., 2019). On the contrary, data-driven models, grounded in empirical data, offer flexibility in analysis, ranging from macro to micro perspectives, depending on the research focus. Examples include probabilistic ship domain analyses (Zhang et al., 2023), machine learning-based risk predictions (Murray and Perera, 2021), and the integration of AIS and accident data for specific accident type risk analysis (e.g., W. Zhu et al., 2023; Zhang et al., 2022a,b). However, these singular perspectives, whether macro or micro, tend to provide limited theoretical guidance. Thus, a comprehensive approach that

### Table 1
Semi-empirical models for maritime traffic safety evaluation.

<table>
<thead>
<tr>
<th>Models</th>
<th>Advantages</th>
<th>Deficiencies</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTA</td>
<td>The model exhibits a notable simplicity in its form, coupled with a convenient calculation process.</td>
<td>Qualitative evaluation methods lack the capability to quantify risks effectively.</td>
<td>Antos and Guedes Soares (2006)</td>
</tr>
<tr>
<td>ETA</td>
<td>The model facilitates an incisive analysis of specific factors influencing maritime accidents.</td>
<td>These methods are subject to the influence of the decision-maker’s subjective judgment, potentially compromising the accuracy of risk evaluations.</td>
<td>Celik et al. (2010)</td>
</tr>
<tr>
<td>Bayesian Networks</td>
<td>Modeling the inherent uncertainty within a model is crucial for effectively addressing incomplete or noisy data.</td>
<td>Predominantly relying on accident data, these approaches often suffer from inaccuracies and incompleteness.</td>
<td>Hannington and Kujala (2012)</td>
</tr>
<tr>
<td>Failure Mode Effect Analysis</td>
<td>Accounting for all conceivable failure modes is essential to ensure the system’s reliability and safety.</td>
<td>Risk assessments are generally conducted from a macro perspective, limiting their ability to provide targeted guidance for policymakers.</td>
<td>Sharma et al. (2008)</td>
</tr>
<tr>
<td>Fuzzy-Logic</td>
<td>This approach addresses many challenges associated with imprecise and uncertain data, accommodating for limited information and the evolving nature of available knowledge.</td>
<td>Determine the ship domain based on the knowledge of the relevant navigator.</td>
<td>Balmat et al. (2011)</td>
</tr>
<tr>
<td>System-Theoretic Accident Model and Processes</td>
<td>Analysis of past accidents enables the design of safer systems, proactively preventing accidents during the system development phase.</td>
<td>The use of CSSD data derived from AIS big data enables effective estimation without relying on the expert knowledge required by the fuzzy ship domain.</td>
<td>素质教育 (2019)</td>
</tr>
<tr>
<td>Fuzzy domain</td>
<td>The ship domain is determined based on the expertise and knowledge of the relevant navigator.</td>
<td>We propose a collision risk index estimation model utilizing a support vector machine.</td>
<td>Liu et al. (2020)</td>
</tr>
<tr>
<td>Fuzzy quaternion ship domain</td>
<td>Utilizing concepts of uncertainty and fuzzy information, this method defines the fuzzy boundaries of the ship’s domain and estimates the possibility of spatial collisions.</td>
<td>A regional collision risk prediction framework is developed based on real-time AIS data, employing the Shapley method to define each ship’s risk contribution.</td>
<td>Liu et al. (2020)</td>
</tr>
</tbody>
</table>

### Table 2
Data-driven models for maritime traffic safety evaluation.

<table>
<thead>
<tr>
<th>Models</th>
<th>Advantages</th>
<th>Deficiencies</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirically-calibrated ship domain</td>
<td>This model integrates features of both analytical and empirical ship domains.</td>
<td>The limited availability of datasets poses challenges in distinguishing factors influencing the shape and size of the ship domain.</td>
<td>Wang and Chin (2016)</td>
</tr>
<tr>
<td>Probabilistic ship domain</td>
<td>The use of CSSD data derived from AIS big data enables effective estimation without relying on the expert knowledge required by the fuzzy ship domain.</td>
<td>/</td>
<td>Zhang et al. (2019)</td>
</tr>
<tr>
<td>Ship collision risk using AIS</td>
<td>A kinematics feature-based ship collision ranking operator is introduced for measuring real-time traffic conflicts.</td>
<td>The impact of ship type on quantifying ship collision risk has not been incorporated.</td>
<td>Liu et al. (2023)</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>We propose a collision risk index estimation model utilizing a support vector machine.</td>
<td>The model’s effectiveness may diminish with the adoption of higher values for the collision risk index.</td>
<td>Gang et al. (2016)</td>
</tr>
<tr>
<td>Collision risk prediction</td>
<td>A regional collision risk prediction framework is developed based on real-time AIS data, employing the Shapley method to define each ship’s risk contribution.</td>
<td>The accuracy of predictions is impacted by the clustering effect within the data.</td>
<td>Liu et al. (2020)</td>
</tr>
<tr>
<td>Real-time ship collision risk evaluation</td>
<td>The MPC index is proposed to enhance the precision of collision risk perception using ship dimension information.</td>
<td>The computation of the MPCA relies on the assumption that the ship’s heading remains constant.</td>
<td>Zhao et al. (2021); Zhao and Fu (2021)</td>
</tr>
<tr>
<td>Dynamic multi-period maritime accident assessment</td>
<td>Dynamic multi-period maritime accident prediction models are established by combining accident data with AIS data, facilitating susceptibility evaluation.</td>
<td>Risk assessments are confined to existing data, with a notable absence of prospective studies on future accident susceptibility.</td>
<td>Zhu et al. (2023)</td>
</tr>
<tr>
<td>Information entropy approach with AIS</td>
<td>Information entropy is utilized to develop indices for the quantitative measurement of collision risks.</td>
<td>The results’ validity has not been corroborated with additional collision data.</td>
<td>Feng et al. (2022)</td>
</tr>
<tr>
<td>Risk factors and navigation accidents</td>
<td>The combination of AIS data and accident data aids in identifying indicators and factors associated with navigational accidents.</td>
<td>The scope of the study is limited to individual ships, lacking consideration of regional risk factors.</td>
<td>Aalberg et al. (2022)</td>
</tr>
<tr>
<td>Spatial-temporal analysis method</td>
<td>Analyzing the spatial-temporal characteristics during ship encounters contributes to Key ship characteristics, such as relative speed, have not been factored into the analysis.</td>
<td>Key ship characteristics, such as relative speed, have not been factored into the analysis.</td>
<td>Zhu et al. (2023)</td>
</tr>
</tbody>
</table>

(continued on next page)
amalgamates both macro and micro perspectives is imperative. Integrating the spatial-temporal dynamics of ship traffic with extensive maritime accident data allows for a richer, multi-dimensional understanding of maritime risks. This integrated perspective is pivotal in offering decision-makers a broader and more informed basis for developing comprehensive maritime traffic safety strategies and policies.

1.2. Contributions

Ship traffic within inland waterways, notably in the Nanjing section of the Yangtze River, exhibits a higher level of complexity, when compared to open sea environments (e.g., Zhang et al., 2013; Zhang et al., 2022a,b). The intricate nature of these waterways is characterized by obstructive structures and challenging navigational conditions. Obstacles such as cross-river bridges, the sinuous nature of waterways, and the uneven distribution of land masses contribute to the increased navigational demands on ships. It is observed that traffic conditions can vary significantly between adjacent areas and waterways (Wang et al., 2014). While previous research has explored traffic characteristics of the Yangtze River (Yu et al., 2019), there remains a gap in understanding the correlation between spatial-temporal traffic characteristics and ship navigation safety, especially in the Nanjing section (see section 1.1).

Consequently, this paper presents a novel big data analytics approach to examine the spatial-temporal dynamics of ship traffic in inland waterways. The methodology synergizes raster data, spatial autocorrelation, and Moran’s index with AIS data for a quantitative analysis of maritime traffic. The focus lies on employing the spatial autocorrelation model to capture and illustrate the spatial-temporal evolution of traffic conditions in these complex waterways. This enables a detailed visual representation of the spatial distribution and temporal dynamics of ship traffic in the Nanjing section of the Yangtze River. Additionally, this study integrates real accident data from this area to delve deeper into the impact of spatial-temporal dynamics on ship routing safety and to evaluate the overall safety of maritime traffic in inland waters. The key contributions of this study include:

- The study introduces an innovative big data analytics methodology, integrating raster data with a spatial autocorrelation model. This method is pivotal in quantitatively assessing the spatial-temporal dynamics of ship traffic in inland waters, thereby enriching our understanding of maritime traffic characteristics in these complex environments.
- By combining spatial-temporal dynamics of ship traffic with real accident data, the study provides a nuanced evaluation of maritime traffic risk. It identifies high-risk areas and elucidates the mechanisms of accident evolution in these zones. This analysis is instrumental in providing empirical data to support strategies aimed at reducing maritime traffic accidents.
- This research provides invaluable findings for stakeholders, including maritime authorities, shipping companies, shipowners, and crew members. By offering detailed insights into risk levels, accident causation in high-incidence areas, and the spatial-temporal dynamics of ship traffic near accident sites, the study provides a comprehensive understanding of the spatial correlation between ship traffic dynamics and maritime accidents. This information is critical in aiding these stakeholders to implement effective preventative measures, thereby significantly reducing maritime accidents and associated economic losses.

The paper begins with a brief literature review in section 1, followed by a detailed description of the research methodology in section 2. Section 3 analyzes the flow and speed characteristics of ship traffic, leading to a discussion in section 4, and concluding remarks are presented in section 5.

2. Framework for the analysis of ship traffic spatial-temporal dynamics

The analytical framework devised for analyzing the spatial-temporal dynamics of ship traffic and maritime accidents encompasses two primary components, as illustrated in Fig. 2. This framework is systematically segmented into two pivotal steps.

Step (i): Construction of a spatial-temporal dynamic dataset. This initial step involves compiling a comprehensive dataset that captures both dynamic (such as ship speed, position, and course) and static (including ship type, length, and width) information. The process incorporates techniques like data anomaly detection, drift detection, parking spot detection, and course anomaly detection to refine the raw AIS data. Subsequently, the study areas are demarcated based on the topographical characteristics of the waters under study, enabling a more focused investigation of the spatial-temporal dynamics of ship traffic.

Step (ii): Analysis of spatial-temporal distribution characteristics. Using the dataset from step (i), this phase explores the spatial-temporal patterns in ship trajectory density and average speed. The AIS data is rasterized, followed by the application of a spatial auto-correlation model and Moran’s index. This facilitates a detailed analysis of the spatial correlation between ship average speed and trajectory density. The insights gleaned from this analysis contribute to a deeper understanding of the spatial-temporal distribution of ship traffic flow characteristics. Furthermore, this step is essential in examining the relationship between ship characteristics’ spatial-temporal dynamics and maritime traffic safety, ultimately aiding in the evaluation of maritime traffic safety.

2.1. Step i: data preprocessing

Given that AIS data are susceptible to errors during collection, transmission, and reception phases, rigorous data processing is essential to mitigate potential noise and inaccuracies. Echoing the approach of Zhang et al. (2018), who introduced a multi-state ship trajectory reconstruction model, this study implements targeted data preprocessing techniques to address specific issues encountered in the data. The data preprocessing flow is illustrated in Fig. 3.

The methods applied are delineated as follows.

(i) Data anomaly detection: This process entails identifying and rectifying various data exceptions. Anomalies are flagged in scenarios such as when the Maritime Mobile Service Identity (MMSI) call sign does not conform to the standard 9-digit format or contains non-numeric characters. Similarly, ship longitude readings exceeding 180°, latitude readings beyond 90°, negative ship speeds, and course values surpassing 360° are all categorized as anomalies and subjected to correction.

(ii) Parking spot detection: This method identifies and removes data points that indicate prolonged inactivity of ships, thus eliminating redundant information. To enhance the accuracy and relevance of the data, points indicating prolonged ship inactivity, evidenced by static longitude and latitude coordinates, are identified and excluded from the dataset, as per the methodology proposed by Wang et al. (2020).

(iii) Drift detection: Drift in AIS data represents significant deviations from a ship’s expected trajectory, often resulting from
transmission or positioning errors. This study involves the identification and elimination of such abnormal data points, ensuring that the ship’s trajectory is accurately represented and optimized for subsequent analysis.

(iv) Course anomaly detection: The paper determines the abnormality of the ship’s course by evaluating RoT. The RoT is defined and utilized to identify abnormal course variations. The RoT is calculated as:

\[
RoT = \frac{|H_{tn} - H_{tn-1}|}{t_n - t_{n-1}}
\]

where RoT is calculated by comparing the ship heading \(H_{tn}\) at time \(t_n\) with the ship heading \(H_{tn-1}\) at time \(t_{n-1}\). Filipiak et al. (2018) present methods for anomaly detection in the maritime domain tested in a traditional and big data setting, which detects a sharp change of course (over 90°). It is assumed that a ship should not change course so quickly, and if this happens, it may be interpreted as loitering, indicating a significant and abrupt change in the ship heading at \(t_{n-1}\). Such data points are not suitable for extracting accurate ship trajectories and therefore are excluded from the analysis (Hoque and Sharma, 2020). For the Course anomalies, this paper will combine the linear interpolation method to repair the trajectories of the anomalies. The linear interpolation can be calculated using the following equation (2):

\[
l_i = l_{i-1} + (t - (t - 1)) \times \left(\frac{l_{i+1} - l_{i-1}}{(t+1) - (t-1)}\right)
\]

where \(t\) is the current time of the course anomaly point, \(l_{i-1}\) and \(l_{i+1}\) represent the longitude (or latitude) data of two adjacent time points \(t - 1\) and \(t + 1\) of the course anomaly point, respectively. \(l_i\) is the result (longitude or latitude) of linear interpolation calculation at time \(t\).

2.2. Step ii: characteristic analysis methods

In this section, the key components of the methodology used in this study, including raster data, ship trajectory density, ship average speed, global spatial autocorrelation, and local spatial autocorrelation, are described in detail.

2.2.1. Raster data

Raster and vector data models represent two pivotal approaches for structuring spatial data. Raster data effectively consolidates complex datasets into discrete spatial entities, streamlining analytical methods and enhancing scalability for large data volumes, as explained by Rawson et al. (2022). Pertinently, within the ambit of maritime traffic flow studies, raster data proffers distinct advantages over its vector counterpart.

- Enhanced spatial resolution: Raster data, represented as a matrix of grids or pixels, offers superior spatial resolution. This detail allows for precise analysis of traffic dynamics and understanding of maritime traffic flow’s spatial patterns.
- Optimal adaptability: Raster data’s grid-based structure aligns well with the diverse scales in maritime traffic analysis, making it suitable for complex maritime scenarios.
- Intuitive visualization and efficient data processing: Raster data’s segmental nature aids in intuitive representation and understanding...
of maritime traffic patterns, enhancing the clarity of research findings. Its structure also allows for streamlined data processing.

It merits emphasis that in a raster dataset, each pixel epitomizes a consistently dimensioned area, as illustrated in Fig. 4. The granularity of the grid cells is adjustable, tailored to the requisite level of detail for representing targeted maritime elements. Finer cells engender a more nuanced and detailed representation within the grid, albeit at the expense of increased pixel count, thereby extending processing durations and amplifying storage requisites, as expounded by Brisaboa et al. (2020). Consequently, to best capture the interaction between ship trajectory density and average speed distribution in nearby traffic flow dynamics, a grid cell size of 200 × 200 was chosen. This methodology ensures a robust and accurate portrayal of maritime traffic characteristics and adjacent flow attributes.

2.2.2. Ship trajectory density

There are two commonly used methods for analyzing ship trajectory density. The first approach involves generating ship trajectory lines using the latitude and longitude information from the AIS data. These lines are then superimposed onto a grid. The density of ship trajectories within each grid cell is determined by counting the number of trajectories that intersect with the cell. For example, as depicted in Fig. 5 (a), area C, devoid of any ship trajectories, has a density of 0, while area C, intersected by two ship trajectories, has a density of 2. While straightforward, this method has notable limitations, such as not accounting for trajectory lengths within each cell, as highlighted by Liu et al. (2022a,b). A case in point is that areas C and D in Fig. 5 (a) both have two trajectories, but the lengths of the trajectories in area C far exceed those in area D. To address the limitations of the first method, this paper adopts an enhanced technique, termed traffic linear density, illustrated in Fig. 5 (b). Here, the trajectory density in each grid cell is calculated by summing the lengths of all ship trajectories within that cell. This method provides a more nuanced measurement by considering the total trajectory length, thereby offering a more accurate representation of ship traffic density within each grid area. The two trajectory densities are calculated as:

\[
\text{Traj\_den}_i = \frac{\text{Traj\_num}_i}{\text{Area}} (3)
\]

\[
\text{Traj\_den}_i = \frac{\sum \text{Traj\_len}_i}{\text{Area}} (4)
\]

where \(\text{Traj\_den}_i\) is the density of ship trajectories in the \(i^{th}\) grid cell. \(\text{Traj\_num}_i\) is the number of ship trajectories in the \(i^{th}\) grid cell, \(\text{Area}\) is the size of grid cell. \(\text{Traj\_len}_i\) is the length of the \(j^{th}\) ship trajectories in the \(i^{th}\) grid cell.

2.2.3. Ship average speed

This paper primarily emphasizes the analysis of the average speed of ships. The speed dispersion on the grid refers to the standard deviation of the ship average speed calculated from all ship trajectories within a grid cell. The average speed of ship \(i\) can be calculated using the following equation (5):

\[
\bar{v}_i = \frac{\sum_{j=1}^{n_i} \|D_{i+1} - D_i\|}{t_n - t_1} (5)
\]

where \(\bar{v}_i\) is the average speed of ship \(i\) in the grid, \(t_n - t_1\) is the sailing time difference of the ship, \(n_i\) is the number of trajectory points of the ship \(i\) in the grid. \(\|D_{i+1} - D_i\|\) is the absolute difference between the distance travelled by ships at different periods.
Assuming that \( N_s \) ships are passing through the grid cell, the ship average speed at the grid cell is calculated by equation (6):

\[
E(\tau_i) = \frac{\sum_{i=1}^{N_s} \tau_i}{N_s}
\]

where \( E(\tau_i) \) represents the average speed of all ships in the grid cell, \( N_s \) is the number of ships passing through the grid (Wang et al., 2020).

2.2.4. Global spatial autocorrelation

In marine traffic feature research, Moran’s index stands out for its marked advantages over other spatial analysis tools. This index offers more than just a quantitative assessment of spatial correlation, it adeptly identifies areas of high aggregation (hotspots) or dispersion. Such capabilities enable a profound exploration of spatial distribution trends in specific geographical phenomena. Moreover, the versatility of Moran’s index extends to temporal comparisons, equipping researchers with the tools to analyze spatial-temporal variations in marine traffic features. This, in turn, contributes to a more holistic understanding and facilitates informed decision-making. Therefore, this paper selects Moran’s index as the spatial analysis tool, and the subsequent sections will provide a detailed exposition of its specific applications, encompassing spatial weight matrix, global spatial autocorrelation, and the following subsection on local spatial autocorrelation.

The water areas are divided into \( N \) small cells. The spatial weight matrix is calculated as:

\[
W = [w_{ij}]_{N \times N} = \begin{bmatrix}
0 & w_{12} & \cdots & w_{1N} \\
w_{21} & 0 & \cdots & w_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
w_{N1} & w_{N2} & \cdots & 0
\end{bmatrix}
\]

where \( w_{ij} \) represents the spatial weight between grid cells \( i \) and \( j \). Moran’s index can be quantified using either the adjacency or the distance criterion. In the adjacency criterion, spatial cells that are connected are assigned a value of 1, indicating proximity, while unconnected cells receive a value of 0, denoting no direct spatial relationship. Conversely, the distance criterion operates on a range basis, where cells within a specified distance are given a value of 1, and those beyond this range are assigned a value of 0. This distinction allows for flexibility in defining spatial relationships based on different research needs or geographic contexts.

The global spatial Moran’s index, as described by Jackson et al. (2010), functions as a correlation coefficient. It quantitatively assesses the overall degree of spatial autocorrelation for a given set of attribute values in conjunction with their spatial positions. This index is particularly useful in providing a comprehensive view of spatial autocorrelation within a dataset. The calculation equation of the global spatial Moran’s index is shown in equation (8):

\[
I = \frac{N}{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}
\]

where \( N \) is the number of grid cells, \( x_i \), \( x_j \) are the attribute value involved in the grid cell in this study (ship average speed and trajectory density), \( \bar{x} \) is the mean value of all observations, \( w_{ij} \) is the spatial weight between grid cells \( i \) and \( j \).

The global spatial Moran’s index ranges from \(-1\) to \(1\) and reflects the spatial distribution pattern of the research object. If \( I > 0 \), the traffic states are spatially positively correlated, with values close to 1 indicating a strong relationship. If \( I < 0 \), the traffic states are spatially negatively correlated, with values close to \(-1\) indicating a close relationship. If \( I = 0 \), the traffic states are not spatially related.

In addition, the P-value represents the probability that the observed spatial pattern is created by some random process. The Z-score represents the multiple of standard deviation, and the Z-score mainly reflects the degree of dispersion of the data set. The four distributions of P-values and Z-values are as follows (Rong et al., 2021).

- If the P-value is \(< 0.01\) and the Z-score is \(> 2.58\), it is considered that 99% of the elements are clustered.
- If the P-value is \(< 0.01\) and the Z-score is \(< -2.58\), 99% of the elements are considered to be discrete.
- If the P-value is \(< 0.01\), but the Z-score is \(< 2.58\), the pattern may be the result of a random spatial process.
- If the P-value is small and the absolute value of the Z-score is large, it indicates that the observed spatial pattern cannot be the result of a random process.

The degree of significance of spatial autocorrelation is illustrated in Fig. 6.

In the analysis of Moran’s index, the P-value and Z-score are critical metrics that warrant careful consideration. As elucidated by Xu et al. (2023), a small P-value indicates that the observed spatial distribution is unlikely to be a result of random chance, suggesting significant spatial patterns or correlations in the dataset. This suggests a meaningful spatial pattern or correlation within the data set under study. Conversely, the Z-score, as highlighted by Chen (2021), functions as a measure of dispersion. It is calculated as the deviation from the mean, expressed in terms of standard deviations. The Z-score offers insight into the degree
of dispersion or clustering of the research subjects. Together, these two metrics provide a comprehensive understanding of the spatial characteristics being analyzed. While the P-value assesses the significance of the spatial pattern, the Z-score quantifies the degree of this spatial deviation. Their combined interpretation is essential for validating the spatial autocorrelation results obtained through Moran’s Index, thereby ensuring robust and reliable conclusions in spatial analysis studies.

2.2.5. Local spatial autocorrelation

The application of local spatial autocorrelation analysis is pivotal in understanding the spatial dynamics within specific regions. The importance of this analysis is twofold:

Uncovering hidden local spatial autocorrelation: In scenarios where global spatial autocorrelation is absent, local spatial autocorrelation analysis is instrumental in identifying potential hidden autocorrelations at a local level. Conversely, in cases where global spatial autocorrelation is present, this analysis assists in evaluating spatial heterogeneity, thereby offering a more nuanced understanding of the spatial distribution.

Identifying spatial outliers or influential points: This method is effective in pinpointing spatial outliers or areas of strong influence that diverge from the trends indicated by global spatial autocorrelation. For example, even when global spatial autocorrelation analysis suggests a positive correlation, local analysis may reveal pockets of negative spatial autocorrelation, thus highlighting spatial anomalies.

In the context of spatial statistical analysis, both the global and local Moran’s indexes are extensively utilized, as noted by Cheng et al. (2012). In this study, the local Moran’s index is specifically applied to analyze ship trajectory density and average speed within the designated study area. Unlike the global Moran’s index, which is confined to a range of −1 to 1, the local Moran’s index does not adhere to such constraints. This characteristic allows for a more flexible and detailed examination of spatial patterns at the local level, providing valuable insights into the spatial dynamics of ship traffic conditions. The calculation equations for the local Moran’s index \( I_l \) is provided in equation (9):

\[
I_l = \frac{(n-1)(x_i - \bar{x}) \sum_{j=1}^{n} w_{ij} (x_j - \bar{x})}{\sum_{j=1}^{n} (x_j - \bar{x})^2}
\]

(9)

where \( x_i \), \( x_j \) are the attribute value involved in the grid cell in this study (ship average speed and trajectory density), \( \bar{x} \) is the mean value of all observations, \( w_{ij} \) is the spatial weight between grid cells \( i \) and \( j \), and \( n \) is equal to the total number of grid cells. Unlike the global Moran’s index, the local Moran’s index does not have to be in the range [-1,1]. If \( I_l > 0 \), the traffic state is positively correlated with its spatial local neighbors with larger values indicating a stronger relationship. If \( I_l < 0 \), the traffic state is negatively correlated with its local neighbors, with smaller values indicating a stronger relationship. If \( I_l = 0 \), then \( x_i \) has no relationship with its local neighbors.

3. Case study

The Nanjing section of the Yangtze River holds a pivotal role in China’s inland waterway transportation network, distinguished by its substantial traffic flow and the highest freight volume compared to other inland riverine cities (Zhao et al., 2019). This segment, spanning a length of 150 km and encompassing 332 wharf berths, presents a complex navigational landscape. It features varied water topographies such as cross-river bridges, curved and narrow waterways, and interspersed land areas, which collectively heighten navigational challenges and amplify the potential for water traffic accidents, as detailed in Appendix A (Cai et al., 2021).

In analyzing this critical section, the study area is stratified into three distinct regions based on topographic characteristics: Nanjing 1, with five continental land domains; Nanjing 2, including a significant tributary; and Nanjing 3, devoid of continental domains or major tributaries. Temporal segmentation is also applied, dividing the day into three intervals: early morning (0:00–8:00), daytime (8:00–16:00), and evening (16:00–24:00). This segmentation facilitates a thorough examination of traffic characteristics across different geographies and timeframes.

Applying the methodology delineated in section 2.2, the traffic dynamics within the Nanjing section are meticulously explored. The analysis focuses on varying geographical locations and temporal periods, providing a comprehensive overview of the maritime traffic patterns in this strategically important segment of the Yangtze River.

3.1. Data source

3.1.1. Experimental dataset description

To conduct an empirical analysis of the traffic characteristics in the Nanjing section of the Yangtze River (Table 3), a comprehensive dataset of AIS data was compiled. This dataset, encompassing the period from 2019 to 2021, includes a total of 7,701,914 records from ships operating in the area. The data are organized into two distinct databases. The first database houses dynamic AIS data, encompassing various parameters such as MMSI, ship position, time, speed, course, draft, and other pertinent information from over 1400 ships (Table 4). The second database is dedicated to static AIS data, containing details like ship MMSI, type, length, and beam.

To streamline the analysis of this extensive dataset, the waters within the Nanjing section were divided into three distinct research areas: Nanjing 1, Nanjing 2, and Nanjing 3, as shown in Table 3. The delineation of these areas is illustrated in Fig. 7. Additionally, to assess the influence of ship traffic characteristics on navigational safety, maritime accident data were sourced from the Nanjing Maritime Bureau. Emphasizing the relevance and accuracy of the data, the study focused on recent years, selecting accident records from 2019 to 2021. This decision was informed by the understanding that older accident data might not adequately reflect current traffic scenarios and conditions.

Table 3

<table>
<thead>
<tr>
<th>Study area</th>
<th>Coordinate 1</th>
<th>Coordinate 2</th>
<th>Coordinate 3</th>
<th>Coordinate 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lat/Lon (N/° E’)</td>
<td>Lat/Lon (N/° E’)</td>
<td>Lat/Lon (N/° E’)</td>
<td>Lat/Lon (N/° E’)</td>
<td></td>
</tr>
<tr>
<td>Nanjing 1</td>
<td>31.76/118.52</td>
<td>32.27/118.52</td>
<td>32.27/118.67</td>
<td>31.76/118.67</td>
</tr>
<tr>
<td>Nanjing 2</td>
<td>31.76/118.67</td>
<td>32.27/118.67</td>
<td>32.27/119.01</td>
<td>31.76/119.01</td>
</tr>
<tr>
<td>Nanjing 3</td>
<td>31.76/119.01</td>
<td>32.27/119.01</td>
<td>32.27/119.15</td>
<td>31.76/119.15</td>
</tr>
</tbody>
</table>
Consequently, a total of 42 recent maritime accident records were incorporated into the study, providing a contemporary perspective on the safety challenges within this complex inland waterway (Table 5).

### 3.1.2. AIS data preprocessing

This section delineates the preprocessing of AIS data, with a particular focus on addressing noise and anomalies in the original ship trajectories. As depicted in Fig. 8(a), notable anomalies are evident in the ship course, including steering angles exceeding 90°. Such irregularities can significantly compromise the accuracy of traffic flow data statistics within the study area. To address this issue, points where the course appears abnormal are removed, and the resulting gaps are filled using linear interpolation. This approach results in the formation of a revised typical trajectory, as exemplified in Fig. 8 (b). Additionally, ‘drift’ in the data, characterized by abnormal points significantly diverging from the typical trajectory, is addressed. These deviations, often a result of transmission or positioning errors, are evident in Fig. 9 (a), where the ship trajectory erroneously veers towards the shore. To correct this, drift points are excised, and the affected trajectory is segmented into one or more coherent parts, as illustrated in Fig. 9 (b).

For the detection of data anomalies and parking spots, specific thresholds are established based on the normality of AIS data and the kinematic characteristics of the ships. This step is crucial in filtering out erroneous data, indicating that 9.83% of the total 7,701,914 records have been removed. Fig. 10 displays a comparative analysis between selected excerpts of AIS data, both processed and unprocessed, to demonstrate the effectiveness of the preprocessing methods used in enhancing data reliability for the experiments conducted in this study.

### Table 5

**Accident characteristic data sample.**

<table>
<thead>
<tr>
<th>Number</th>
<th>Ship type</th>
<th>Accident type</th>
<th>Lat</th>
<th>Lon</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bulk cargo ship</td>
<td>Collision</td>
<td>32.1771</td>
<td>119.0002</td>
<td>4:30-5:00</td>
</tr>
<tr>
<td>2</td>
<td>Liquefied gas carrier</td>
<td>Fire/ explosion</td>
<td>32.2413</td>
<td>119.1452</td>
<td>21:20-22:30</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>41</td>
<td>Oil tanker</td>
<td>Collision</td>
<td>31.9707</td>
<td>118.6435</td>
<td>4:30-5:00</td>
</tr>
<tr>
<td>42</td>
<td>Bulk cargo ship</td>
<td>Collision</td>
<td>31.9708</td>
<td>118.6434</td>
<td>2:00-3:00</td>
</tr>
</tbody>
</table>

### 3.2. Spatial-temporal dynamic of ship trajectory density

#### 3.2.1. Ship trajectory density analysis

The trajectory density of ships varies due to temporal and spatial factors. Fig. 11 shows that the trajectory density is relatively high in the northern area of Zimu Island in Nanjing 1. The trajectory density in this region generally exceeds 2.641. This is because the area accommodates both upbound and downbound ships, and the channel here is relatively narrow, leading to a higher trajectory density. In Nanjing 1, the number of grids with high ship trajectory density significantly increases during the period from 8:00 to 16:00, with maximum ship trajectory density ranging from 2.743 to 3.135. In Nanjing 2, encompassing the Jiajiang tributaries and the Yangtze River, Fig. 12 shows that the ship trajectory density in the Jiajiang River is lower than in the main channel of Yangtze River. Nevertheless, the trajectory density near Zhongshan dock on the Yangtze River channel is higher. The highest ship trajectory density occurs from 8:00 to 16:00, ranging between 0.449 and 0.959. Nanjing 3 features a dock on its north bank where most commercial ships berth. Consequently, Fig. 13 illustrates that the trajectory density in the near-shore waters of the northern part of Nanjing 3 is high, with the majority of areas having a density exceeding 2.750. In Nanjing 3, the maximum ship trajectory density is 3.072 between 16:00 to 24:00 and slightly higher at 3.096 from 0:00 to 8:00, a difference of just 0.024.

#### 3.2.2. Global spatial autocorrelation analysis of ship trajectory density

To analyze the correlation between spatial location and ship trajectory density, the paper utilizes Moran’s index and global autocorrelation methods (see section 2.2.4). Table 6 presents the Moran’s index, Z-score, and P-value for each period in the Nanjing section. Table 6 indicates that the P-value for all regions and each period is 0.000, and the Z-score for all regions and each period is significantly greater than 2.580. These findings indicate a 99% probability that the spatial distribution pattern of ship trajectory density is not randomly distributed, with a probability of random distribution being less than 1% (Rong et al., 2021). The lowest Z-score, observed at 9.194 during 0:00–8:00 in Nanjing 2, suggests a relatively lower spatial correlation of trajectory density in this area compared to others during the same period. Moreover, Moran’s index for all regions is positive, indicating a positive spatial correlation in ship trajectory density within the Nanjing section. The strongest positive correlation is observed during the period of 0:00–8:00 in Nanjing 1, with a Moran’s index of 0.629, while the weakest positive...
A correlation is found during the period of 0:00–8:00 in Nanjing 2, with a Moran’s index of 0.264. Notably, the correlation in Nanjing 2 exhibits more variability across different periods, with a difference of 0.329 between the largest and smallest Moran’s index values. Based on the results of this study, the variations in Moran’s index, P-value, and Z-score across different periods underscore the significant impact of spatial location on ship trajectory density, highlighting the importance of considering spatial variations in maritime traffic safety analysis.

3.2.3. Local spatial autocorrelation analysis of ship trajectory density

In general, local spatial autocorrelation models are effective in identifying potential locations of specific local correlations that may be hidden, facilitating the evaluation of spatial heterogeneity. As shown in Fig. 8, the waters near Nanjing Banqiao exhibit the highest local Moran’s index, indicating a high local spatial correlation of trajectory density, possibly due to the unique geographical features or concentrated shipping activities. Grids closer to the main route, demonstrate higher spatial correlation of trajectory density, while those farther away show lower correlation. The local spatial correlation of ship trajectory density in the region shows a large variation. The difference of maximum Moran’s index is 0.1011 during 0:00 to 8:00 and 8:00 to 16:00, reflecting significant shifts in local spatial correlations during these times. Fig. 14 demonstrates that areas with a high spatial correlation of trajectory density in Nanjing 2 are primarily located near the Nanjing Qixiashan Yangtze River Bridge and around Zhongshan dock. The local spatial correlation of ship trajectory density in this area varies greatly.

Fig. 8. Schematic diagram of course anomaly detection, (a) ship trajectory with anomalous course, (b) ship trajectory without anomalous course.

Fig. 9. Schematic diagram of drift detection, (a) ship trajectory with drift points, (b) ship trajectory without drift points. In particular, the red dots denote drift points.

Fig. 10. Schematic diagram of data preprocessing, (a) unprocessed ship trajectories, (b) processed ship trajectories.
significantly between 0:00 to 8:00 and 8:00 to 24:00, which may be related to the operation time of the ferry at Zhongshan Ferry Terminal. In Fig. 16, waters with high trajectory density correlation in each period of Nanjing 3 are distributed along the north bank and the main route. This pattern is influenced by the presence of a dock on the north shore, leading to high trajectory density and subsequent spatial correlation in these areas.

3.3. Spatial-temporal dynamic of ship average speed

3.3.1. Analysis of ship average speed

Based on the ship average speed (see section 2.2.3), this study investigates the spatial distribution characteristics of ship average speed in Nanjing section of the Yangtze River in different periods. Regarding spatial distribution, Fig. 17 illustrates that the average speed of upbound ships in Nanjing 1 is notably lower compared to downbound ships. The main route of the Yangtze River exhibits ship average speed ranging from 7 to 10 knots, dropping to lower averages near the Nanjing Da Shengguan Yangtze River Bridge. In Fig. 18, the average speed of ships in the Jiajiang tributaries of Nanjing 2 is significantly lower than that of ships on the main line of the Yangtze River. Ships in the tributaries predominantly maintain speed between 1 and 4 knots, while ships on the main route typically range from 3 to 8 knots. Additionally, within the main line of the Yangtze River, upbound ships exhibit lower average speed than downbound ships. Fig. 19 demonstrates that the traffic flow in Nanjing 3 is relatively straightforward, with consistent speed distributions in each period. In this region, the average speed of upbound ships is lower than that of downbound ships, upbound ships average speed are between 0 and 3 knots, while downbound ships range from 5 to 10 knots. In terms of time distribution, the variation in ship average speed is more pronounced across different periods than that of ship trajectory density. In three regions (Nanjing 1, Nanjing 2, and Nanjing 3), the period from 8:00 to 16:00 generally exhibits higher ship average speed, predominantly falling within the range of 6–10 knots. Additionally, in all three regions and across all time periods, only a small number of ships have an average speed exceeding 10 knots.

3.3.2. Analysis of global spatial autocorrelation of ship average speed

In this study, the ship average speed is somewhat similar to the ship trajectory density in terms of P-value, but the Moran’s index and Z-value are more variable. Table 7 shows that the P-value is consistently 0.000 for all regions and periods, with Z-scores significantly exceeding 2.580 in each case. These findings indicate a 99% probability that the spatial distribution pattern of ship average speed is not randomly distributed, with the likelihood of random distribution being less than 1% (Rong et al., 2021; Zhang et al., 2024). Notably, the lowest Z-score is observed from 0:00 to 8:00 in Nanjing 2, indicating that the ship average speed in this area exhibits a relatively weak spatial correlation compared to other areas during that period. Moreover, the Moran’s index for all regions is positive, there is a clear indication of positive spatial correlation in ship average speed throughout the Nanjing section. The strongest positive correlation is observed during the 8:00–16:00 in Nanjing 1, with a Moran’s index of 0.564. Conversely, the weakest positive correlation is observed during the 16:00–24:00 in Nanjing 2, with a Moran’s index of 0.116. Notably, the correlation within Nanjing 1 shows greater variability across different periods, with a difference of 0.283 between the largest and smallest Moran’s index. Overall, the correlation analysis of Moran’s index, p-values and z-scores shows that spatial-temporal variations have an effect on the ship average speed.
3.3.3. Analysis of local spatial autocorrelation of ship average speed

To further examine the impact of spatial-temporal variation on ship average speed, the study utilizes the local correlation method (see section 2.2.5). The local Moran’s index of ship average speed in each period within the Nanjing section is calculated. Fig. 20 reveals that the high local correlation of ship average speed in Nanjing 1 is primarily concentrated in the southern part of Zimu Island, with a maximum Moran’s index of 0.078. The overall local spatial correlation of ship average speed is higher in Nanjing 1 from 8:00 to 16:00 compared to the other two periods. In Fig. 21, a more evenly distributed pattern of local correlations in the average speed of ships in Nanjing 2, with Moran’s index predominantly ranging from 0.002 to 0.003. However, there are
still limited regions with high local correlations beyond this range. For example, the Nanjing 2 region also exhibits high local spatial correlation in ship average speed from 8:00 to 16:00, with a maximum Moran’s index of 0.122. Fig. 22 indicates that the downstream routes and the vicinity of the docks in Nanjing 3 exhibit a higher local correlation in average ship speed, with the Moran’s index ranging between 0.006 and 0.008. This suggests a strong spatial correlation in average ship speeds in this area. The local autocorrelation distribution pattern of the ship average speed in this region is similar in each period.

Overall, the analysis of local spatial autocorrelation of ship average speed reveals specific patterns within different regions and time periods.

These findings contribute to a better understanding of the relationship between ship average speed and spatial-temporal factors within the Nanjing section.

4. Discussions

This section analyzes the results of the case study and the impact of traffic characteristics on ship navigational safety, and summarizes the implications of this study for maritime management science.

4.1. Influence of ship traffic characteristics on ship safety

The empirical study indicates that the spatial-temporal autocorrelation in maritime transportation networks is dynamic and spatially heterogeneous. A significant correlation exists between maritime traffic accidents and these spatial-temporal dynamics. The relationship between accident distribution and ship traffic dynamics, including average speed and trajectory density, is detailed in Appendix B.

Maritime traffic accidents in Nanjing 1 primarily occur in areas with high ship trajectory density along the Yangtze River. Some accidents also happen in docks or near-shore waters. In Nanjing 2, accidents take place in the tributaries of the JiaJiang River and the main channel of the Yangtze River, but most of them occur in areas with high ship trajectory density on the main and curved waterways prone to accidents. Additionally, accidents also occur in the waters with high ship trajectory density in Nanjing 3.

Table 6
Global spatial autocorrelation index of ship trajectory density in the Nanjing section.

<table>
<thead>
<tr>
<th>Area</th>
<th>Index</th>
<th>0:00–8:00</th>
<th>8:00–16:00</th>
<th>16:00–24:00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moran’s index</td>
<td>0.629</td>
<td>0.621</td>
<td>0.530</td>
<td></td>
</tr>
<tr>
<td>Z-score</td>
<td>19.992</td>
<td>19.708</td>
<td>15.476</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.000 (&lt;10^{-2})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moran’s index</td>
<td>0.264</td>
<td>0.593</td>
<td>0.544</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.000 (&lt;10^{-2})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moran’s index</td>
<td>0.412</td>
<td>0.413</td>
<td>0.422</td>
<td></td>
</tr>
<tr>
<td>Z-score</td>
<td>10.466</td>
<td>10.482</td>
<td>10.713</td>
<td></td>
</tr>
<tr>
<td>P-value</td>
<td>0.000 (&lt;10^{-2})</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 14. Local Moran’s index of ship trajectory density in Nanjing 1.

Fig. 15. Local Moran’s index of ship trajectory density in Nanjing 2.

Fig. 16. Local Moran’s index of ship trajectory density in Nanjing 3.

Table 6 indicates that the spatial-temporal autocorrelation in maritime transportation networks is dynamic and spatially heterogeneous. A significant correlation exists between maritime traffic accidents and these spatial-temporal dynamics. The relationship between accident distribution and ship traffic dynamics, including average speed and trajectory density, is detailed in Appendix B.

Maritime traffic accidents in Nanjing 1 primarily occur in areas with high ship trajectory density along the Yangtze River. Some accidents also happen in docks or near-shore waters. In Nanjing 2, accidents take place in the tributaries of the JiaJiang River and the main channel of the Yangtze River, but most of them occur in areas with high ship trajectory density on the main and curved waterways prone to accidents. Additionally, accidents also occur in the waters with high ship trajectory density in Nanjing 3.

These findings contribute to a better understanding of the relationship between ship average speed and spatial-temporal factors within the Nanjing section.
Fig. 16. Local Moran’s index of ship trajectory density in Nanjing 3.

Fig. 17. Ship average speed distribution in Nanjing 1.

Fig. 18. Ship average speed distribution in Nanjing 2.

Fig. 19. Ship average speed distribution in Nanjing 3.
The location of maritime traffic accidents in Nanjing 1 is less influenced by speed. These accidents mainly occur in the waters around the Nanjing Yangtze River Bridge, suggesting that fixed obstacles in the river impact safe navigation. In Nanjing 2, accidents are distributed around the Nanjing Yangtze River Bridge, Nanjing Ba Guazhou Yangtze River Bridge, and the curved narrow waterway area. In Nanjing 3, accidents are concentrated in the waters along the Yangtze River where ship average speed are high. The majority of maritime traffic accidents in the Nanjing section are classified as general-grade accidents, with ship collisions being the predominant accident type. Collisions can cause significant damage to assets and result in personal injuries (Zhang et al., 2021). This underscores the impact of ship trajectory density and ship average speed on ship navigation safety, leading to a high incidence of ship collisions and unnecessary economic losses (Yang et al., 2023).

Table 7
Global spatial autocorrelation index of ship average speed in the Nanjing section.

<table>
<thead>
<tr>
<th>Area</th>
<th>Moran’s index</th>
<th>Z-score</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanjing 1</td>
<td>0.528</td>
<td>16.745</td>
<td>0.000 (&lt;10⁻²)</td>
</tr>
<tr>
<td></td>
<td>0.564</td>
<td>17.855</td>
<td>8.357</td>
</tr>
<tr>
<td></td>
<td>0.281</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nanjing 2</td>
<td>0.146</td>
<td>5.192</td>
<td>0.000 (&lt;10⁻²)</td>
</tr>
<tr>
<td></td>
<td>0.161</td>
<td>5.866</td>
<td>5.241</td>
</tr>
<tr>
<td></td>
<td>0.116</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nanjing 3</td>
<td>0.313</td>
<td>7.956</td>
<td>0.000 (&lt;10⁻²)</td>
</tr>
<tr>
<td></td>
<td>0.326</td>
<td>8.276</td>
<td>6.944</td>
</tr>
<tr>
<td></td>
<td>0.271</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Q. Ma et al.

Fig. 20. Local Moran’s index of the ship average speed in Nanjing 1.

Fig. 21. Local Moran’s index of the ship average speed in Nanjing 2.

Fig. 22. Local Moran’s index of the ship average speed in Nanjing 3.
Recent research on maritime accidents, which pose significant threats to life and property, focuses on a range of aspects. Studies like Zhang et al. (2019), Wang and Chin (2016), and Liu et al. (2023) examine collision risks within ship domains, but their scope is limited to individual ships without encompassing regional risk assessments. Others, such as Gang et al. (2016), Liu et al. (2020), and Kim and Lim (2022), employ machine learning for predicting accidents and assessing risks, contributing to future risk management. However, these studies often neglect causal analysis. Semi-empirical approaches including Bayesian Networks, Fuzzy Logic, and Fault Tree analysis have been utilized to explore accident causes, as seen in works by Hänninen and Kujala (2012), Balmat et al. (2011), and Chen et al. (2022a, b), but they are heavily dependent on expert knowledge. Addressing these gaps, incorporating spatial-temporal dynamics of ship traffic with accident data into maritime risk evaluation offers a comprehensive view, enabling regional risk assessment and revealing the underlying mechanisms of accident evolution (Q. Zhu et al., 2023). The utility of AIS data and accident records in maritime accident risk evaluation is increasingly recognized (Aalberg et al., 2022).

In the era of big data, developing a method for maritime accident risk evaluation that utilizes diverse datasets and integrates ship spatial-temporal dynamics is of great value. This study presents a method that effectively visualizes the spatial-temporal distribution of ship traffic in areas prone to accidents, including metrics such as average ship speed and trajectory density. This approach enhances understanding of spatial correlations in specific regions, assisting in probing the internal mechanisms of accident evolution. Furthermore, it allows for the identification of risk areas within the study waters through the spatial-temporal correlation of ship traffic features with accidents. It is generally observed that areas with high ship traffic density and rapid average speeds exhibit a higher frequency of accidents. However, discrepancies exist, as some high-risk areas may not experience accidents, and vice versa, underscoring that risk and accidents do not always share a direct causal link (Cucinotta et al., 2017). While accidents retain a degree of randomness, a heightened risk level typically suggests a greater likelihood of accidents occurring. This understanding forms the crux of our research into maritime traffic safety evaluation.

4.2. Managerial insights from this study

Maritime risk evaluation is a critical area of focus for scholars and policymakers, as it plays a vital role in ensuring maritime traffic safety and offering data support and technical guidance. For this reason, this study is summarized as follows.

(1) Implications for ocean and coastal management or governance.

By analyzing the spatial-temporal dynamics of ship traffic and factors affecting maritime traffic accidents, the results of the study can guide the ocean and coastal management sector to take preventive measures against maritime traffic accidents and to strengthen water traffic control. Government agencies such as the Maritime Safety Administration and the Ministry of Transport can use the results of the study to propose targeted preventive measures. For example, using the ship routing to delineate warning zones based on high-risk areas for accidents, to ensure that ships can navigate safely after entering high-risk areas, and to increase the vigilance of maritime navigators. Furthermore, reasonable adjustment of ship sailing patterns for areas with high spatial correlation, such as bridge areas and narrow waterways, to minimize multi-ship encounters. Logistics and shipping companies can leverage the study’s findings to fine-tune their scheduling, strategically avoiding periods with high ship density (Chen et al., 2023). This approach not only minimizes the risk of maritime accidents but also curtails time and economic losses. Therefore, their risk management could become much more scientific and cost-effective (Jiang et al., 2023).

(2) Managerial implications for maritime policymakers.

The results of the study offer valuable insights for maritime decision-makers, such as shipping companies, ports, and docks. Initially, the findings can be used to prioritize areas at risk of maritime accidents both temporally and spatially, so that they can have a more detailed understanding of the risk situation throughout the navigable waters. Specifically, shipping companies can use it to plan routes and select docks, and ports can use it to optimize ship berthing and unberthing patterns. For instance, the presence of a commercial dock in the northern part of the Nanjing 3 results in a high density of ships throughout the all day, and consider increasing the number of ship berths to improve service levels and optimize traffic flow patterns in the area.

(3) Managerial implications for maritime navigators.

The results of this research offer crucial data-driven insights for maritime navigators. Specifically, in areas characterized by high risk, dense ship traffic, and elevated speeds, it is imperative for navigators to heighten their alertness to mitigate the risk of maritime accidents. Such incidents not only carry the risk of penalties but, in extreme cases, can also result in loss of life. By adopting a more vigilant and precautionary approach in these critical areas, the likelihood of maritime incidents can be substantially reduced. This proactive stance is pivotal in enhancing overall maritime safety.

4.3. Limitations and future works

The methodology delineated in this paper, integrating spatial-temporal attributes of maritime traffic with accident datasets, constitutes a pivotal framework for executing exhaustive safety evaluations in inland waterways. However, it is imperative to recognize certain constraints and potential avenues for refinement in this research.

- Incomplete maritime traffic representation in AIS data: The AIS data used in this research does not fully capture all maritime traffic in inland waterways. Notably, smaller ships, especially those under 300 tons that typically do not use AIS transponders, are excluded from the analysis. This omission results in a gap in the complete representation of maritime traffic in these areas.
- Lack of standardized quantification scales: The experimental cases in this study faced challenges with standardized quantification due to significant variations in traffic characteristics at different times and regions with relatively small Moran’s index values. Implementing standardized scales could result in a uniform color representation in certain areas, which may reduce visibility and readability, potentially masking intricate spatial and temporal traffic patterns.
- Need for in-depth spatial-temporal analysis: The evaluation of spatial-temporal distribution is preliminary. Future research should focus on developing predictive models for spatial-temporal distribution, providing more profound insights into the patterns and trends of maritime traffic accidents. Such advancements would greatly enhance the effectiveness of maritime traffic risk assessments.

Acknowledging and addressing these aspects are essential to enhance the comprehension and effectiveness of maritime traffic safety assessments in maritime traffic.

5. Conclusions

With the acceleration of economic growth and the burgeoning demand for goods transportation, maritime transportation has emerged as an increasingly vital mode of conveyance. This growth has precipitated a surge in both the number and diversity of ships navigating inland waterways, thereby intensifying the complexities of maritime traffic
management and elevating navigational risks. In response to these challenges, this paper introduces a novel method for evaluating maritime traffic risk, grounded in the spatial-temporal dynamics of ship movements and leveraging AIS data. This method integrates raster data, spatial autocorrelation, and Moran’s index to examine the spatial-temporal distribution of maritime traffic via AIS data, offering visual insights into the distribution and temporal dynamics of ship traffic in the study area. Additionally, by analyzing real accident data, the study explores the influence of ship traffic’s spatial-temporal dynamics on navigational safety, providing a quantitative evaluation of maritime traffic safety to identify risk areas and understand the accident evolution mechanism. The paper employs AIS and accident data from the Nanjing section of the Yangtze River (2019–2021) to demonstrate the method’s effectiveness, yielding significant findings, as outlined below.

- The utilization of AIS data for rasterizing maritime traffic is shown to be an effective approach in evaluating the characteristics of ship traffic.
- Analysis of maritime traffic and ship average speed distribution in the Nanjing section identifies areas with high trajectory density, as well as hotspot characterized by high ship average speed and disturbance.
- The spatial autocorrelation and Moran’s index model were used to examine the relationship between traffic characteristics and spatial-temporal variation. It was found that the P-value was 0.000 (<10^{-3}) and the Z-score consistently exceeded 2.580 for each period and region (see Tables 6 and 7). The results indicate that the spatial patterns of ship trajectory density and ship average speed in Nanjing waters are not randomly distributed. Instead, there are regions with significant spatial correlations of ship trajectory density and ship average speed, influenced by temporal and locational changes.
- The study identifies a correlation between ship traffic characteristics and traffic safety. Hotspot areas for ship average speed and ship trajectory density correspond closely with regions where maritime accidents frequently occur. Maritime traffic accidents predominantly take place in bridge areas, narrow waterways, curved sections of the river and the vicinity of the docks. These areas typically have complex traffic flows and higher requirements for the driver and pilot seamanship, which is also the root cause of the occurrence of accidents.

The methodology proposed in this study enhances the accuracy of maritime traffic risk assessment, addressing the challenge posed by increasingly complex maritime traffic and the resultant high frequency of accidents. Authorities, policymakers, shipowners and crews can benefit from the methodology. The results of the study can assist them in effectively preventing maritime traffic accidents. Additionally, the results of this study provide valuable insights for ensuring safe ship navigation in the Nanjing section and serve as a reference for future spatial analyses of maritime accidents.

CRediT authorship contribution statement

**Quandang Ma:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Huan Tang:** Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Cong Liu:** Writing – review & editing, Visualization, Conceptualization. **Mingyang Zhang:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. **Dingze Zhang:** Writing – review & editing, Visualization, Software, Data curation, Conceptualization. **Zhao Liu:** Writing – review & editing, Software, Resources, Funding acquisition, Data curation, Conceptualization. **Liye Zhang:** Writing – review & editing, Methodology, 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

Data will be made available on request.

Acknowledgments

This study is supported by the National Natural Science Foundation of China, China (Grant No. 52171351) and the Natural Science Foundation of Hubei Province (2022J0089).

Appendix A. The number of accidents in the Yangtze River from 2011 to 2020, distributed by type of accident and jurisdiction (Sui et al., 2023)

Table A1

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Type of maritime accidents</th>
<th>Contact/Collision</th>
<th>Grounding</th>
<th>Capsize/Foundering</th>
<th>Fire/Exploded</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanjing</td>
<td></td>
<td>298</td>
<td>22</td>
<td>12</td>
<td>10</td>
<td>8</td>
<td>350</td>
</tr>
<tr>
<td>Nantong</td>
<td></td>
<td>268</td>
<td>26</td>
<td>15</td>
<td>8</td>
<td>29</td>
<td>346</td>
</tr>
<tr>
<td>Zhongjiang</td>
<td></td>
<td>214</td>
<td>7</td>
<td>19</td>
<td>7</td>
<td>6</td>
<td>253</td>
</tr>
<tr>
<td>Taizhou</td>
<td></td>
<td>148</td>
<td>2</td>
<td>20</td>
<td>7</td>
<td>21</td>
<td>198</td>
</tr>
<tr>
<td>Taizhou</td>
<td></td>
<td>155</td>
<td>5</td>
<td>14</td>
<td>2</td>
<td>15</td>
<td>191</td>
</tr>
<tr>
<td>Chongqing</td>
<td></td>
<td>50</td>
<td>88</td>
<td>22</td>
<td>18</td>
<td>11</td>
<td>189</td>
</tr>
<tr>
<td>Huangshi</td>
<td></td>
<td>122</td>
<td>34</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>179</td>
</tr>
<tr>
<td>Yichang</td>
<td></td>
<td>36</td>
<td>83</td>
<td>9</td>
<td>23</td>
<td>6</td>
<td>157</td>
</tr>
<tr>
<td>Zhangjiagang</td>
<td></td>
<td>127</td>
<td>3</td>
<td>13</td>
<td>6</td>
<td>7</td>
<td>156</td>
</tr>
<tr>
<td>Wuhu</td>
<td></td>
<td>83</td>
<td>27</td>
<td>18</td>
<td>21</td>
<td>6</td>
<td>155</td>
</tr>
<tr>
<td>Yangzhou</td>
<td></td>
<td>130</td>
<td>2</td>
<td>6</td>
<td>14</td>
<td>4</td>
<td>141</td>
</tr>
<tr>
<td>Wuhan</td>
<td></td>
<td>69</td>
<td>38</td>
<td>16</td>
<td>14</td>
<td>4</td>
<td>126</td>
</tr>
<tr>
<td>Jiujiang</td>
<td></td>
<td>86</td>
<td>23</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>126</td>
</tr>
<tr>
<td>Jiangyin</td>
<td></td>
<td>77</td>
<td>0</td>
<td>21</td>
<td>1</td>
<td>3</td>
<td>102</td>
</tr>
<tr>
<td>Yueyang</td>
<td></td>
<td>56</td>
<td>29</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>101</td>
</tr>
<tr>
<td>Jingzhou</td>
<td></td>
<td>39</td>
<td>36</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>89</td>
</tr>
</tbody>
</table>

(continued on next page)
Table A1 (continued)

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Type of maritime accidents</th>
<th>Contact/Collision</th>
<th>Grounding</th>
<th>Capsize/Foundering</th>
<th>Fire/Exploded</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anqing</td>
<td></td>
<td>54</td>
<td>12</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>88</td>
</tr>
<tr>
<td>Changzhou</td>
<td></td>
<td>74</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>88</td>
</tr>
<tr>
<td>Changshu</td>
<td></td>
<td>54</td>
<td>5</td>
<td>18</td>
<td>4</td>
<td>3</td>
<td>84</td>
</tr>
<tr>
<td>Sanxia</td>
<td></td>
<td>19</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>Yibin</td>
<td></td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Luzhou</td>
<td></td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

Appendix B. Distribution of maritime traffic accidents in relation to maritime traffic characteristics in Nanjing 1, Nanjing 2 and Nanjing 3

Fig. B1. Spatial distribution of maritime traffic accidents in Nanjing 1 and ship trajectory density.

Fig. B2. Spatial distribution of maritime traffic accidents in Nanjing 1 and ship average speed.
Fig. B3. Spatial distribution of maritime traffic accidents in Nanjing 2 and ship trajectory density.

Fig. B4. Spatial distribution of maritime traffic accidents in Nanjing 2 and ship average speed.

Fig. B5. Spatial distribution of maritime traffic accidents in Nanjing 3 and ship trajectory density.

Fig. B6. Spatial distribution of maritime traffic accidents in Nanjing 3 and ship average speed.

Appendix C. Nomenclature
**Abbreviation** | **Definition**
--- | ---
AIS | Automatic Identification System
VTS | Vessel Traffic Services
FTA | Fault Tree Analysis
MMSI | Maritime Mobile Service Identity
Lat | Latitude
Lon | Longitude
RoT | Rate of Turning

### Variable Definitions

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_i$</td>
<td>Ship heading at time $t_i$</td>
</tr>
<tr>
<td>$l_i$, $l_{i-1}$</td>
<td>Longitudinal (or latitude) data of two adjacent time points $i - 1$ and $i + 1$</td>
</tr>
<tr>
<td>$l_i$</td>
<td>Result (longitude or latitude) of linear interpolation calculation at time $t_i$</td>
</tr>
<tr>
<td>$T_{ij}$</td>
<td>Density of ship trajectories in the $i$th grid cell</td>
</tr>
<tr>
<td>$T_{ij}^{num}$</td>
<td>Number of ship trajectories in the $i$th grid cell</td>
</tr>
<tr>
<td>$T_{ij}^{len}$</td>
<td>Length of the $i$th ship trajectories in the $j$th grid cell</td>
</tr>
<tr>
<td>Area</td>
<td>Size of grid cell</td>
</tr>
<tr>
<td>$ar{v}_i$</td>
<td>Average speed of ship $i$ in the grid</td>
</tr>
<tr>
<td>$D_{ij}$</td>
<td>Absolute difference between the distance</td>
</tr>
<tr>
<td>$E(T_j)$</td>
<td>Average speed of all ships in the grid cell</td>
</tr>
<tr>
<td>$N_{ij}$</td>
<td>Number of ships passing through the grid</td>
</tr>
<tr>
<td>$W_{ij}$</td>
<td>Spatial weight between grid cells $i$ and $j$</td>
</tr>
<tr>
<td>$W$</td>
<td>Spatial weight matrix</td>
</tr>
<tr>
<td>$\Sigma$</td>
<td>Mean value of all observations</td>
</tr>
<tr>
<td>$x_i$, $x_j$</td>
<td>Attribute value involved in the grid cell in this study (ship average speed and trajectory density)</td>
</tr>
</tbody>
</table>

### References


