Scilight

International Journal of Transportation and Logistics Research
https://www.sciltp.com/journals/ijtlr

Article

Spatio-Temporal Analysis of Ship Collision Risk: Insights from AIS Data in the Bass Strait Waters, Australia

Jinyi Yu, Jiangang Fei, Prashant Bhaskar and Wenming Shi *

Centre for Maritime and Logistics Management, Australian Maritime College, University of Tasmania, Newnham, TAS 7248, Australia
* Correspondence: Wenming.Shi@utas.edu.au

How To Cite: Yu, J.; Fei, J.; Bhaskar, P.; et al. Spatio-Temporal Analysis of Ship Collision Risk: Insights from AIS Data in the Bass Strait Waters, Australia. *International Journal of Transportation and Logistics Research* **2025**, *1*(1), 6. https://doi.org/10.53941/ijtlr.2025.100006

Received: 19 June 2025
Revised: 28 September 2025
Accepted: 28 September 2025
Published: 27 October 2025

Abstract: The ever-increasing volume of maritime freight increases the risk of ship collisions with devastating consequences. This study conducts a hotspot analysis to address this safety concern. Using Automatic Identification System data in the Bass Strait waters, Australia, the main results are as follows: First, the Getis-Ord general Gi* statistic shows that most of the Bass Strait waters have a low collision risk on a monthly basis, which can be classified into five clusters. Second, the spatial hotspot analysis identifies the Sydney-Melbourne route as the major shipping route with the highest collision risk, followed by the Melbourne-West Coast of Tasmania route, and the Melbourne-Devonport route. Third, ship collision risk maps for different time periods visualize the Port of Melbourne and Devonport as high-risk areas due to their persistently high Getis-Ord Gi* statistics. Finally, ship collision risk in the Bass Strait waters shows clear monthly and hourly trends as well as seasonal and day-night variations. These results provide valuable insights for enhancing vessel maneuverability and strategic channel coordination, thereby reducing the likelihood of ship collisions.

Keywords: ship collision risk; Gi* statistic; spatio-temporal analysis; hotspot analysis; collision risk visualization

1. Introduction

As the volume of maritime freight increases, there is a corresponding increase in the use of large-capacity and high-speed vessels to meet the growing demand for maritime transport. This trend potentially increases the risk of ship collisions, especially in ports, coastal waters, and narrow channels, resulting in significant economic losses, environmental damage, waterway blockages, and human casualties [1,2]. Effective vessel traffic management, harbor channel planning and design, and ship collision prediction are critical to reducing ship collision accidents. Accordingly, advanced equipment such as Automatic Radar Plotters (ARPA), Automatic Identification Systems (AIS), and Electronic Chart Display and Information Systems (ECDIS) are used to help maritime traffic monitors assess ship collision risk in real time, thereby facilitating a better understanding of vessel traffic conditions and preparation for potential collisions [3]. Among these equipment and systems, AIS generates vast volumes of data. With the advancement of modern big data technologies, the mining of the AIS-derived data has gained significant attention, enabling researchers to unveil patterns and pathways of ship collisions. This is essential for decision-making and policy development such as the designation of specific shipping routes, the imposition of speed regulations, and infrastructural advancements to accommodate the ever-increasing flow of maritime traffic.

On the other hand, the use of AIS data to investigate vessel behavior and collision risk in busy and complex navigation areas is of great importance for vessel safety management and planning, which can be enhanced by using a spatio-temporal hotspot analysis to identify locations where ship collisions are most likely to occur [4]. In

Copyright: © 2025 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

practice, the spatio-temporal hotspot analysis, facilitated by platforms such as Hadoop and Spark, has been widely employed in maritime studies, as highlighted by [5]. It helps researchers better understand the spatio-temporal characteristics of ship collision risk hotspots, leading to improvements in operational management and maritime safety. Another notable benefit is that the spatio-temporal hotspot analysis based on AIS data mining can efficiently mitigate human errors, which play a predominant role in ship collisions [6]. Despite these benefits, previous studies have paid insufficient attention to the spatio-temporal dynamics of maritime traffic based on AIS big data.

To address the above shortcoming, this study integrates the AIS data mining technique with spatio-temporal hotspot analysis to provide a comprehensive understanding of ship collision risk. As the Bass Strait is a busy and critical shipping route between mainland Australia and Tasmania, this study uses the Bass Strait waters as a case study and contributes to the existing literature in two ways. First, it adds to previous research on maritime safety and risk assessment by providing a more holistic understanding of ship collision risk from a spatio-temporal perspective. Second, the quantification of ship collision risk and the visualization of ship collision risk maps for different time periods contribute to the identification of high-risk areas, thus providing the basis for improving the operational efficiency and safety of maritime transport.

The remainder of this study is structured as follows: Section 2 reviews the assessment and spatio-temporal analysis of ship collision risk. Section 3 presents the methodology including data collection, vessel collision risk index (CRI) calculation, spatio-temporal cell creation, hotspot analysis, and data storage, processing, and visualization. Section 4 reports and discusses the empirical results. Section 5 presents the main conclusions, implications, and future research directions.

2. Literature Review

2.1. Ship Collision Risk Assessment

Due to the devastating consequences of ship collisions, a number of methods have been employed to assess ship collision risk and important results have been obtained. Qu et al. employed a combination of speed analysis, deceleration and acceleration analysis, and ship domain analysis to assess ship collision risk in the Singapore Strait [7]. They argued that since the ship domain refers to the waters in which navigators wish to maintain distance from other vessels, an overlap in ship domains indicates a higher likelihood of collisions. Silveira et al. compared the distance between the two ships with the collision diameter in the Pedersen's model, estimated the collision candidates, and then multiplied the generic causation probability with the corresponding collision type [8]. However, they found that the generic causation probability may not be appropriate for the studied waterways. Zhang et al. developed the Vessel Conflict Ranking Operator (VCRO), which considers factors affecting the complexity of two vessels encountering each other and ranks the severity of the encounters [9]. Zheng et al. employed a support vector machine (SVM) model to address the problem of overlapping ship domains, and quantitatively analyzed ship collision risk [10]. Liu et al. further considered kinematic features and developed a Kinematic Feature-based Vessel Conflict Ranking Operator (KF-VCRO), using Density-based Spatial Clustering of Applications with Noise (DBSCAN) to identify collision hotspots in the Zhoushan waters [11].

2.2. Spatio-Temporal Hotspot Analysis of Ship Collision Risk

Traditionally, the Moran's I is used to measure spatial autocorrelation [12]. The Getis-ord general Gi* statistic is used to identify the hot and cold spots of the dataset [13]. Tsou used geographic information system (GIS) data visualization to display the traffic density graph, speed distribution graph and course over ground (COG) distribution graph, which helps analyze the spatial and temporal distribution of vessel traffic [14]. Although the spatio-temporal dynamics of vessel traffic using AIS big data has not received sufficient attention, the existing findings have still provided valuable insights for conducting this study.

Zhang et al. applied Hadoop and MapReduce big data technologies and used the local Moran's I index to find high-speed/density hotspots in AIS data. They discovered that hotspot areas corresponded well with the spatial distribution of ship collisions. Nonetheless, they only considered the factors of speed and vessel density when assessing hotspot areas [15]. Rong et al. utilized the Getis-Ord Gi* method to identify near collision hotspots with overlapping ship domains as near collision criteria and assessed the correlation between near collision hotspots and traffic characteristics based on spatial co-occurrence methods [5]. They found that vessel speed did not significantly contribute to the occurrence of near collisions in hotspot areas, and that the most influential traffic characteristics in different hotspots were not necessarily the same. This suggests that relying solely on certain traffic characteristics to assess ship collision risk is insufficient, thereby requiring a more comprehensive approach. Moreover, they argued that employing a uniform ship domain overlap criterion to pinpoint collision hotspots across

varied traffic scenarios may lead to inaccuracies. Other relevant studies contribute to the assessment of ship collision risk. For example, some terrestrial traffic accident studies, such as [16,17], use space-time cube analysis to examine trends in vehicle crashes. However, there is a noticeable lack of studies on the temporally changing patterns of ship collision hotspots. By incorporating a temporal dimension into the space-time cube analysis, the evolution of collision hotspots over different time intervals can be examined. This approach is important for understanding how seasonal vessels like fishing crafts affect collision hotspots.

A close inspection of previous studies reveals that they can be enriched in the following ways. First, this study shifts attention from port areas or straits to wider areas of shipping lanes and waterways by selecting the Bass Strait waters as a case study for further assessment of ship collision risk due to its importance and data availability. Second, to enrich the existing quantitative evidence on ship collisions, this study calculates the vessel *CRI* using the Fuzzy Comprehensive Evaluation (FCE) method. Then, the vessel *CRI* can be determined through the dot product of the weight vector and the membership vector. Third, this study takes a spatio-temporal perspective and conducts a comprehensive hotspot analysis based on AIS data to assess the risk of ship collision in the wider Bass Strait waters.

3. Method

3.1. Data Collection and Processing

According to [18], the main data sources used in maritime accident analyses are historical accident data, expert experience data, and AIS data. Compared with AIS data, the first two data sources have obvious limitations. First, risk assessment modeling based only on historical data may not provide a comprehensive understanding of risk, primarily due to the limited amount of maritime accident data [19]. Moreover, high-severity accidents may receive disproportionate attention, leading to overfitting of the model. As a result, accident data may not be the most appropriate choice for studies focusing on specific types of accident risk (e.g., ship collision risk). Second, the heuristic collection of expert data remains influenced by individual expert biases. More complex and comprehensive models may lead to a significant increase in parameters, and expert fatigue may affect the results [20]. Unlike the historical accident data and expert experience data, AIS data are more objective, abundant, detailed, and accurate, and can be processed using various mathematical and statistical methods [21,22]. Therefore, AIS data are used in this study to analyze ship collision risk in the Bass Strait waters.

The Australian Maritime Safety Authority (AMSA) has developed the Craft Tracking System (CTS) to provide free historical AIS data. The collected AIS data are filtered within the Bass Strait waters based on the following geographical boundaries: minimum longitude (X_{min}): 141.490; maximum longitude (X_{max}): 149.327; minimum latitude (Y_{min}): −41.325; and maximum latitude (Y_{max}): −37.759.

The following two steps are taken to process the collected data. First, AMSA's original data format is an ArcGIS geodatabase point feature class file, which requires conversion to a more universally accessible comma-separated values (CSV) format. This conversion not only makes the data accessible to a variety of analytical tools, but also allows for efficient handling of large datasets using Spark's distributed processing capabilities. The second step is to clean the collected data to ensure the integrity and usefulness of the dataset, which includes the following key procedures: (1) Missing value handling: Rows with missing values (NaN or empty fields) are omitted or replaced by imputation techniques; (2) Outlier handling: Outliers are identified and removed, focusing on anomalously high vessel speeds (e.g., speeds exceeding 50 knots) and implausible coordinate values that could bias the analysis; (3) Date format validation: This study rigorously validates the date and time fields, discarding data with incorrect timestamp formats to ensure the accuracy of the temporal data. These fields are then converted into standardized date/time formats to facilitate further analysis and visualization.

3.2. Ship Collision Risk Index Calculation

As mentioned above, the vessel *CRI* is a critical input in collision hotspot analysis. To calculate the vessel CRI, formula-based approaches are most commonly used and are based on mathematical formulae that consider a number of parameters, such as vessel speed, course, size, type, proximity to other vessels, and prevailing sea conditions. The main advantage of these formula-based approaches lies in their precision and consistency. Given the same inputs, they consistently produce the same results, which is useful for comparative analysis. In addition, they are easily automated, allowing the vessel *CRI* to be calculated in real time as conditions change.

As shown in Figure 1, a typical formula-based method requires the following vessel parameters [23]: (1) Position (x,y): This gives the exact position of the vessel in a coordinate system and is essential for proximity calculations with other vessels; (2) Length: The length of a vessel can influence its maneuverability and stopping

distance; (3) Course ϕ: This is the direction in which the vessel is currently moving. It is crucial for predicting the future position of the vessel and potential collision points; (4) Speed V: The vessel's speed over the ground can influence its ability to avoid collisions, especially in congested waters. The time to closest point of approach (*TCPA*) or distance to closest point of approach (*DCPA*) between vessels is then calculated. The processed AIS dataset used for descriptive analysis contains 3,469,925 vessel position reports. Each record includes type, length, beam, draught, speed, course, and UTC timestamp with a typical horizontal positional accuracy of ±10 m [24]. To characterize sample features, Tables 1 and 2 report the summary statistics of variables and the vessel-type composition, respectively.

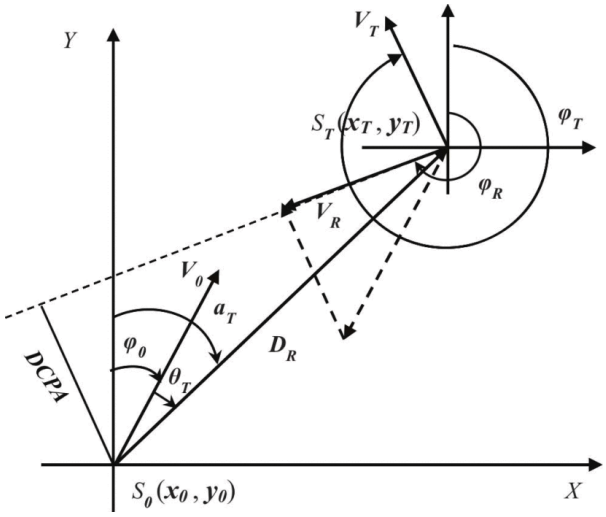

Figure 1. Parameters of encountering vessels.

Table 1. Descriptive statistics of variables.

Variable	Mean	Standard Deviation	Median	Interquartile Range
Speed	3.42	5.95	0.00	5.90
Course	143.08	122.09	134.80	253.10
Length	95.02	95.65	34.00	169.00
Beam	16.50	13.73	11.00	26.00
Draught	4.02	4.37	2.20	7.60

Table 2. Vessel main type frequency and proportion.

Type	Count	Percent (%)
Cargo ship-All	828,290	23.87
Sailing	569,597	16.42
Tug	276,431	7.97
Pleasure craft	246,141	7.09
Tanker-All	198,974	5.73
Fishing	171,960	4.96
Passenger ship	137,025	3.95

For reasons of both practicality and precision, this study employs the model proposed by [25]. Their model calculates the vessel *CRI* using the *FCE* method, which determines the vessel *CRI* by the dot product of two vectors: the weight vector and the membership vector. The model is specified as follows:

$$CRI = W \cdot U = W_{DCPA} \cdot U_{DCPA} + W_{TCPA} \cdot U_{TCPA} + W_D \cdot U_D + W_B \cdot U_{\theta T} + W_K \cdot U_K \quad (1)$$

where W is the weight vector, representing the importance of each factor in the ship collision risk assessment. The membership vector U quantifies the collision risk associated with the encountered vessels, with higher membership values indicating increased collision risk. These membership values are derived from specific membership functions, as described in [25]. As a result, a vessel-level *CRI* for encountering pairs can be computed using Equation (1), where the weight vector W and the membership vector U represent the contributions of *DCPA*, *TCPA*, relative kinematics and headings.

When assessing the *CRI* between vessels, the focus is on vessels that are in an encountering condition. To ensure this, a time-slicing approach is employed to verify that all vessel pairs selected within a given time slice are indeed in such an encounter state. Accurate *CRI* computation requires temporal segmentation of the raw data. Within each time slice, vessel pairs are filtered to confirm their meeting state before the model is applied to calculate the *CRI* for each vessel pair. A critical element of this process is the determination of the meeting condition. To facilitate this, a specified radius is used to identify all vessel pairs within the boundary. According to navigation regulations, vessels may be required to maintain a minimum distance of 1 or 2 nautical miles in certain channels or areas to reduce collisions and other associated risks [26]. These distance requirements are based on factors, such as vessel size, type, and navigational conditions. For example, a large container ship typically requires 1–2 nautical miles to come to a complete stop from full speed, thus necessitating a filter radius of 2 nautical miles for the assessment of vessel pairs.

To ensure that the selected vessel pairs are in a meeting state within the same time slice, it is necessary to interpolate the raw data. As the hotspot analysis is based on monthly, daily, and hourly counts, this study has chosen 10-min intervals for data collection to fill each time slice. The main reason for this choice is that, at an average vessel speed of approximately 12 knots, a ship travels approximately 2 nautical miles in 10 min, which is a perfect match for the chosen filter radius. For the highest temporal granularity, an hour can be divided into 6 intervals, allowing for the precise selection of vessel *CRI* time slices. This approach fully meets the hotspot analysis requirements of this study.

3.3. Spatio-Temporal Cell Creation

In spatial analysis, the integration of the temporal dimension with spatial patterns is crucial. To achieve this, the space-time cube analysis, a three-dimensional geovisualization technique, is employed [27]. This method incorporates both spatial and temporal data within a cube. By adding the *z*-axis (time) to a two-dimensional map with *x*- and *y*-axes representing spatial coordinates, a 3D representation is created. Each event or data point is plotted within this cube based on its spatial location and time of occurrence. While kernel density and hotspot analyses can show where events, such as ship collisions, are concentrated, the space-time cube shows when these concentrations occur. For example, an intersection may be a hotspot for collisions, but the space-time cube can indicate if these collisions are concentrated during rush hours or at weekends. This method also helps identify trends over time. An area may experience a rising collision rate over several months, indicating busier routes or increased traffic. As well as identifying hotspots, the space-time cube can show if events are clustered in time, which is crucial for understanding episodic events or seasonal variations.

To apply the space-time cube analysis, the Bass Strait waters can initially be divided into several small cells. The size of each basic unit cell should be chosen appropriately to identify hotspot areas and should be larger than the largest vessel sailing in the Bass Strait waters. The parameters of the space-time cube are critical to the hotspot analysis and will be adjusted during data analysis. Following the formula method in Figure 1, the collision risk of the vessels sailing in each cell can be considered in the weighting of the cell. The third dimension, time, is then added by stacking the cells vertically along the time axis to create a space-time cube in Figure 2. This helps analyze the trends in hotspots over time.

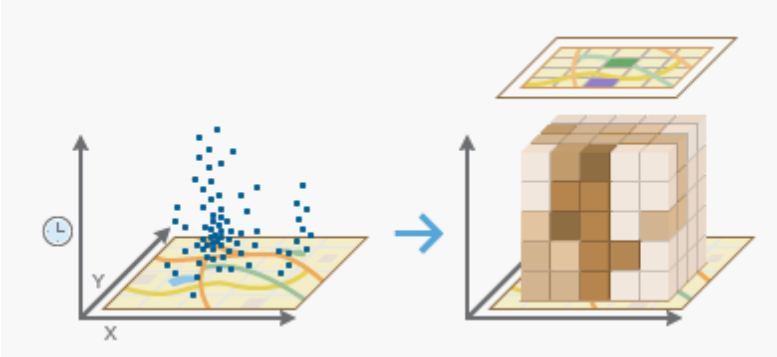

Figure 2. Spatial temporal cells from ArcGIS Pro documentation.

After performing the above process, many small zones are formed in the Bass Strait waters, the number of which depends on the size of the zones, potentially reaching tens of thousands as shown in Figure 2. Based on the parameters chosen to calculate the vessel CRI, an approximate spatial grid cell size of 2 nautical miles was chosen as the side length for each spatial unit. The size of the time interval also affects the number of cells. For example,

in a coarse-grained model of a water area approximately 1 × 1 square degrees in latitude and longitude. If the cell size is 0.001 × 0.001 degrees and the time axis interval is 1 day with data input every 10 s, there will be 8,640,000,000 cells. For each cell, it is necessary to calculate the weights of all its neighbors and determine their Gi* scores which will be introduced in the following section. This study uses the Spark parallel computing engine to make this process feasible and scalable.

3.4. Hotspot Analysis Using the Getis-Ord General Gi* *Statistic*

According to [13], hotspot analysis using the Getis-Ord general Gi* statistic (hereafter the Gi* statistic) is an effective technique for identifying areas of high concentration of features or events. This technique examines the spatial relationships of features to determine whether the observed spatial configuration is due to random events or whether it shows a statistically significant trend. The Gi* statistic compares the local mean (the average value for a location and its neighbors) with the global mean (the average value for all locations). If they are significantly different, the location is identified as a hotspot or coldspot. A hotspot is a location with a high value surrounded by other high value locations, while a coldspot is a location with a low value surrounded by other low value locations. The Gi* statistic is calculated using Equation (2).

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{[n \sum_{j=1}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2]}{n-1}}} \quad (2)$$

where i is the current cell for calculating collision risk, while j is one of the neighboring cells of i. $w_{i,j}$ is the weighting value of the distance between i and j, indicating how j will influence i. x_j is the attribute of j, showing the sum of all risk indices for all vessels in cell j.

To transition from vessel-level risk to cell-level risk, within each time slice, all pairwise CRIs whose geometric mid-points fall into spatial cell j are aggregated. In Equation (2), the cell attribute is thus defined as x_j = **Σ** {encounters in cell$_j$} CRI. This approach makes Gi* a clustering test on aggregated encounter risk rather than merely on traffic density, thereby aligning the two stages of the method. As a result, the Gi* statistic for each cell can be calculated, which yields a z-score and serves as an indicator of collision risk for that cell. Moreover, the z-score contrasts the local mean (a cell and its neighbors) to the global mean across all cells. Positive z indicates a hotspot (high values surrounded by high values), while a negative z indicates a cold spot. Under standard assumptions, the z-scores are approximately normal, that is, $|z| \geq 1.96$ and $|z| \geq 2.58$ correspond to 5% and 1% two-sided significance, respectively. In this study, larger positive z-scores mark intense clusters of encounter risk.

Having obtained the Gi* statistic for each cell at a given time point, this study can choose cells within specific spatial and temporal ranges to assess vessel collision risk and visualize the data accordingly. More detailed analysis can be conducted in the following aspects. (1) Temporal trends: Analyze how the G* statistic evolves over time at specific spatial coordinates. This analysis can provide insight into whether certain areas are experiencing increasing or decreasing collision risk over time; (2) Spatial hotspots: Identify locations (spatial coordinates) with consistently high or low G* statistics. These hotspots can be prioritized for further investigation. Analyzing the distribution of the *CRI* (i.e., Gi* statistic) helps identify areas of increased collision risk. (3) Correlation analysis: Investigate whether there is a correlation between spatial parameters and the G* statistic. Understanding these relationships can shed light on how certain spatial factors may influence risk levels. (4) Temporal spatial distribution: Identify patterns and trends. That is, the spatial distribution of the G* statistic is visualized for each individual month to identify possible patterns or changes, especially seasonal trends or anomalies. For example, consistently high Gi* statistics in certain areas may indicaate underlying structural or operational issues that require attention.

3.5. Data Storage and Processing Framework for Ship Collision Risk Analysis

As mentioned above, data storage and processing are handled by Spark, a big data processing framework known for its scalability and efficiency. Spark's in-memory processing capability significantly increases the speed of data access and analysis, enabling detailed and potentially real-time examination of AIS data. Its flexibility in interacting with different data sources and compatibility with multiple programming languages facilitates seamless integration with vessel management and decision support systems.

The analysis framework is structured as follows: (1) Data storage layer: Data from the CTS system was stored in CSV format in the Hadoop file system. The dataset, covering the period from January 2000 to April 2022, was

organized into files based on a 12-month window; (2) Data ingestion: AIS records were imported into a Spark SQL dataframe from the CSV files; (3) Data parsing: Records were parsed to determine cell coordinates, assigning spatial locations to each data point; (4) Attribute calculation: The CRIs were calculated for each vessel within the defined cells; (5) Neighbor dataframe formulation: A neighboring dataframe was created by merging primary cells with adjacent cells for comprehensive spatial analysis; (6) Parameter calculation: Essential parameters for each cell were calculated for the Gi* statistic calculation; (7) Gi* statistic derivation: The Gi* statistic was calculated for each cell, providing a metric to assess the importance of each cell relative to its neighbors.

3.6. Data Visualization

Effective visualization is crucial for transforming complex statistical metrics into accessible visual formats, making it easier to identify patterns. In this study, visualizations not only illustrate the spatial distribution of ship collision risks, but also provide temporal insight into how these risks evolve over time. Importantly, these visualizations offer practical guidance for vessel operations. For example, a high collision risk map included in the visualizations can help ship officers increase their vigilance during deck watches. By visually depicting areas of increased collision risk, the map enables ship officers to remain more alert and proactive in these critical areas. This increased awareness can lead to improved risk management and accident prevention, thereby enhancing the overall safety of maritime operations.

4. Empirical Results and Discussion

4.1. Distribution of the Gi* *Statistic*

After applying the method described in Section 3, the Gi* statistics are obtained. When the data are segmented by month, the distribution of the Gi* statistics is presented in Figure 3, with the following main results: First, there are a total of 11,507 time-space cells with the Gi* statistics. Most of these values are clustered around zero, with many being negative. This suggests that for a significant part of the dataset, there is no significant spatial clustering of high or low values. The prevalence of negative values indicates that coldspots (areas of low collision risk) are more common than hotspots (areas of high collision risk). Second, the frequency of the Gi* statistics decreases as the value approaches its maximum. The maximum value of 35.946073 is an outlier, significantly deviating from the majority of the dataset. This indicates that there are very few areas with extremely high collision risk. Third, the standard deviation of 2.119664 indicates a moderate dispersion in the data. Additionally, the 75th percentile remains negative, reinforcing the observation that most data points are coldspots. In summary, these results suggest that although there are some areas of increased collision risk, they are relatively rare. Most of the Bass Strait waters have lower collision risk on a monthly basis.

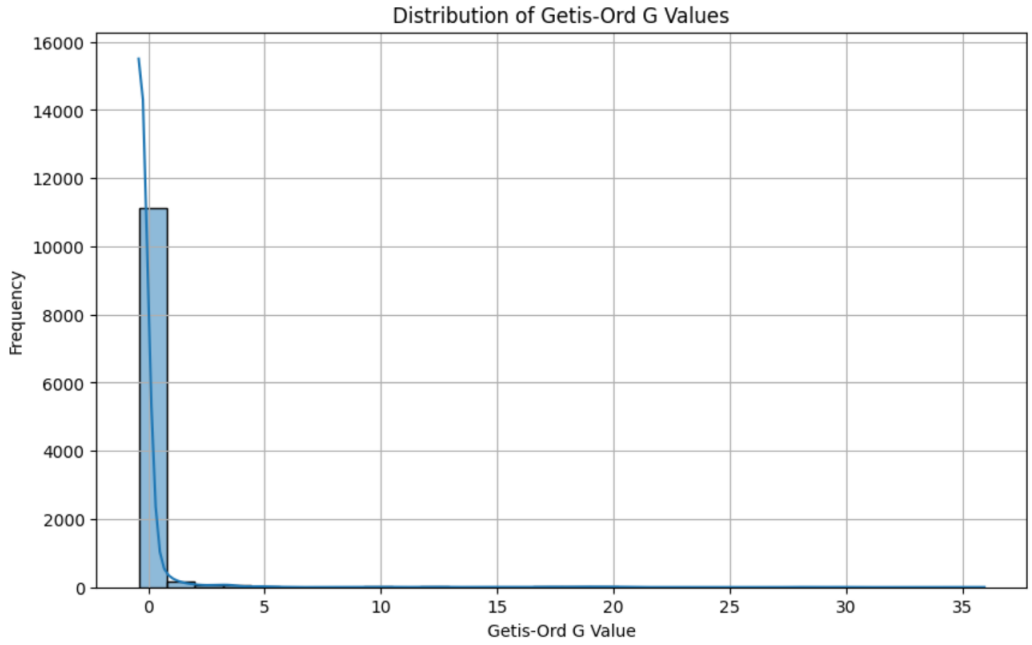

Figure 3. Distribution of the Gi* statistic values.

4.2. Ship Collision Risk along Major Shipping Routes

For the calculated Gi* statistics over geographical coordinates (longitude and latitude), this study applies cluster analysis to identify major shipping routes with similar ship collision characteristics. Technically, cluster analysis groups data points based on common characteristics, segmenting the dataset into meaningful clusters. The K-Means algorithm is commonly used for this purpose, which divides the data into a predefined number of clusters by minimizing the distance between data points and their cluster centers. In this study, the following characteristics are considered for clustering: (1) Spatial coordinates (longitude and latitude): These are essential to determine the exact geographical position of each data point; (2) Temporal coordinates (hour or month): These add a temporal dimension, allowing analysis of patterns over time; (3) The Gi* statistics: These measure local spatial association and highlight areas with particularly high or low values. These characteristics collectively support the clustering process and facilitate the identification of regions with similar collision characteristics.

Using the K-Means algorithm, the dataset has been divided into five predetermined clusters. Figure 4 presents the clustering results, with each color representing one of these different clusters, based on spatial coordinates, temporal coordinates, and the Gi* statistics. (1) Blue cluster: This indicates a high collision risk zone along the maritime corridor between Sydney and Melbourne. This area is likely to have an increased collision risk due to factors such as heavy vessel traffic, environmental conditions, or complex navigation; (2) Orange cluster: This represents the collision risk within the Melbourne port area, including both vessel entry and exit points. The high Gi* statistics here indicate significant spatial clustering, possibly due to the high concentration of vessels in the port area; (3) Green cluster: This covers shipping routes between Melbourne and Tasmania, including major routes to/from Hobart, Burnie, and Devonport. This cluster may reflect well-established maritime routes or frequently travelled routes; (4) Red cluster: This highlights the northeastern route along the Tasmanian coast, where vessels navigate close to islands. This cluster may represent frequently used or strategically important shipping corridors; (5) Purple cluster: This represents routes between Melbourne and Adelaide, indicating another important shipping corridor. The clustering analysis provides valuable insights into vessel movements and identifies potential collision risk areas in the Bass Strait waters.

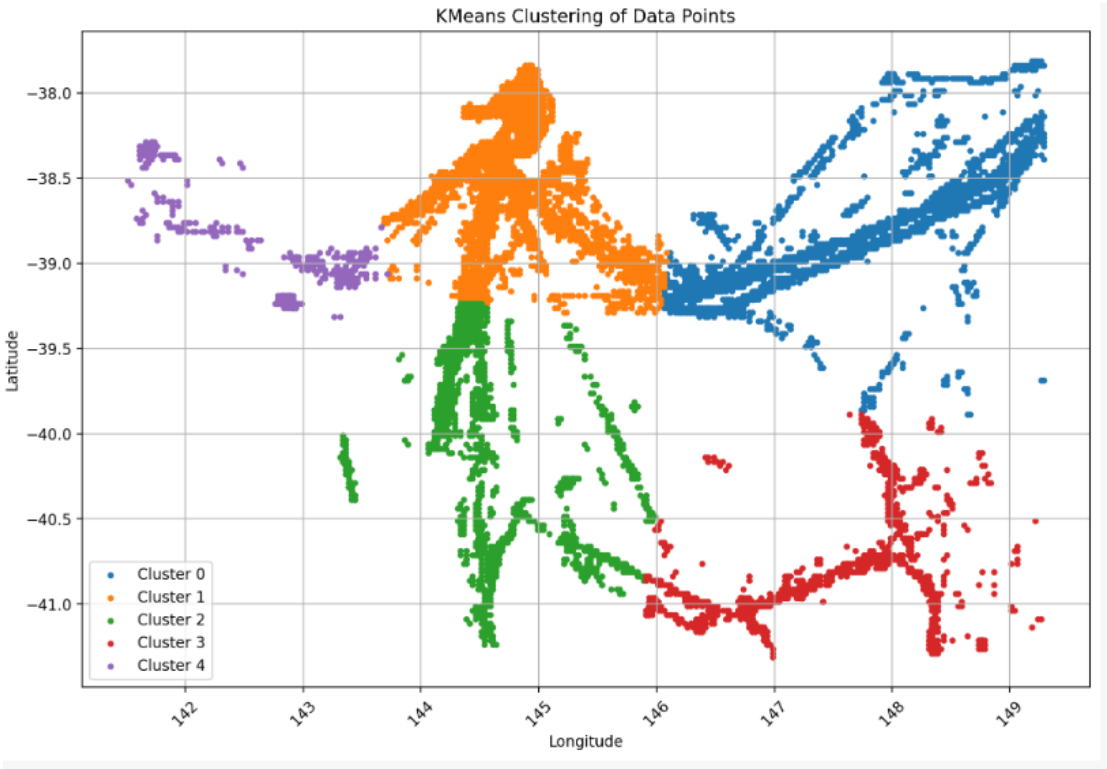

Figure 4. K-Means clustering of data points.

4.3. Spatial Hotspot Analysis Based on Sustained Gi* *Statistics*

Figure 5 visualizes the Gi* statistics, with red dots indicating areas of consistently high values and blue dots indicating areas of consistently low values. This visualization helps identify regions with varying levels of ship collision risk over time. A close inspection of Figure 5 reveals that areas (e.g., Melbourne Port and Devon Port), marked by red dots, have consistently high Gi* statistics. This persistence may be due to factors such as high

traffic volumes, complex shipping routes, or a combination of these and other factors. To mitigate these risks, targeted measures such as specific navigation rules or warning signs could be implemented. These measures aim to provide better guidance to vessels, reduce congestion, and minimize navigational confusion, thereby enhancing maritime safety. Conversely, areas marked by blue dots demonstrate consistently low Gi* statistics, indicating a lower collision risk. Despite the lower risk, it is important to remain vigilant. As unexpected incidents can still occur, continuous monitoring and adaptable regulations are essential to ensure that these areas remain protected from potential ship collisions. Persistently high Gi* areas require targeted interventions, such as operating rules, signage, and patrol allocation to mitigate concentrated encounter risk. In contrast, persistently low Gi* areas serve as baselines that validate the model and guide resource allocation toward routine monitoring rather than intensive control. The clear distinction between regimes makes the analysis actionable, directing mitigation efforts to stable hotspots while confirming where lighter-touch surveillance is sufficient.

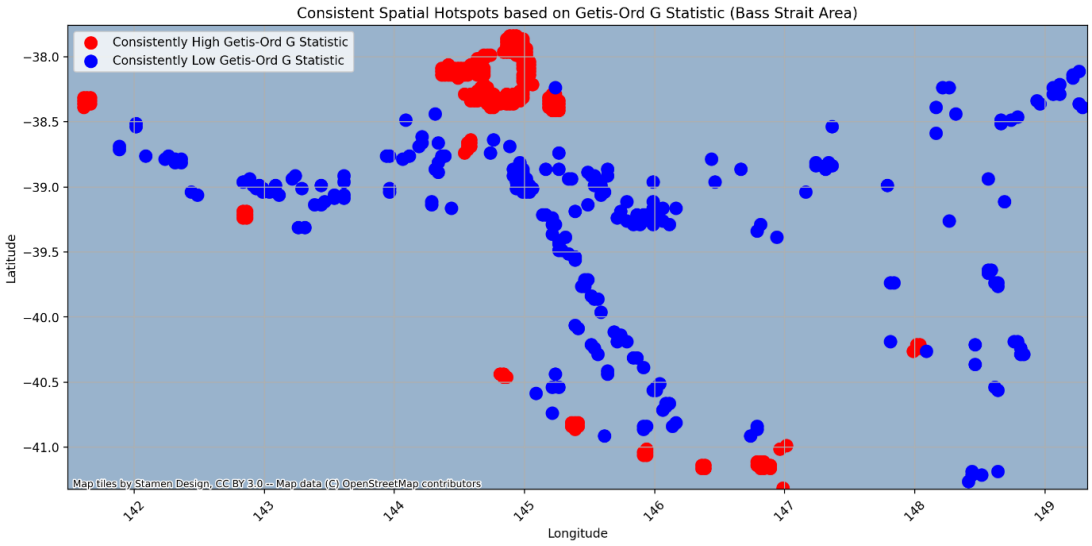

Figure 5. Spatial hotspot analysis.

4.4. Monthly Trend Analysis and Seasonal Variations

This study further examines the average Gi* statistics and plots their monthly trend in Figure 6. It can be seen that the average Gi* statistics are significantly low from December to August, typically showing negative values, but become positive in September, October and November. This indicates seasonal variations in maritime activity in the Bass Strait waters. For example, the average Gi* statistics increase from July to October, indicating an increased collision risk during these months. Conversely, they decrease in November and December, indicating a decreased risk. From January to July, ship collision risk demonstrates more variability. This pattern aligns with the Bass Strait's winter sea conditions (e.g., strong winds, rough seas, reduced visibility) as well as seasonal activity rhythms (e.g., fishing restrictions and returns), which increase encounter complexity and promote the aggregation of *CRI* into spatial clusters [28]. Although the monthly mean generally declines during November–December, a late-December spike in 2022 coincided with the Melbourne–Hobart yacht race (27 December), temporarily increasing encounter densities along the race corridor [29]. These seasonal variations may be influenced by factors such as fishing closures and weather conditions. The weather in the Bass Strait waters, which is notoriously variable, presents challenges such as turbulent seas, strong winds, and reduced visibility from July to October. These adverse conditions may contribute to the observed increase in the Gi* statistics during this period. On the other hand, Figure 7 visualizes the spatial distribution of the average Gi* statistics for each month of 2022, from January (Month 1) to December (Month 12). A 4 × 3 grid of subplots is created, with the color intensity of each data point representing the magnitude of the value. This visualization clearly shows how the Gi* statistics change over time and space. The above monthly trend analysis and seasonal variations provide valuable insight into factors affecting navigational safety in the Bass Strait waters. They highlight the importance of incorporating seasonal considerations into maritime planning, safety protocols, and further research. By doing so, stakeholders can enhance navigational safety, optimize maritime operations, and protect the marine ecosystem.

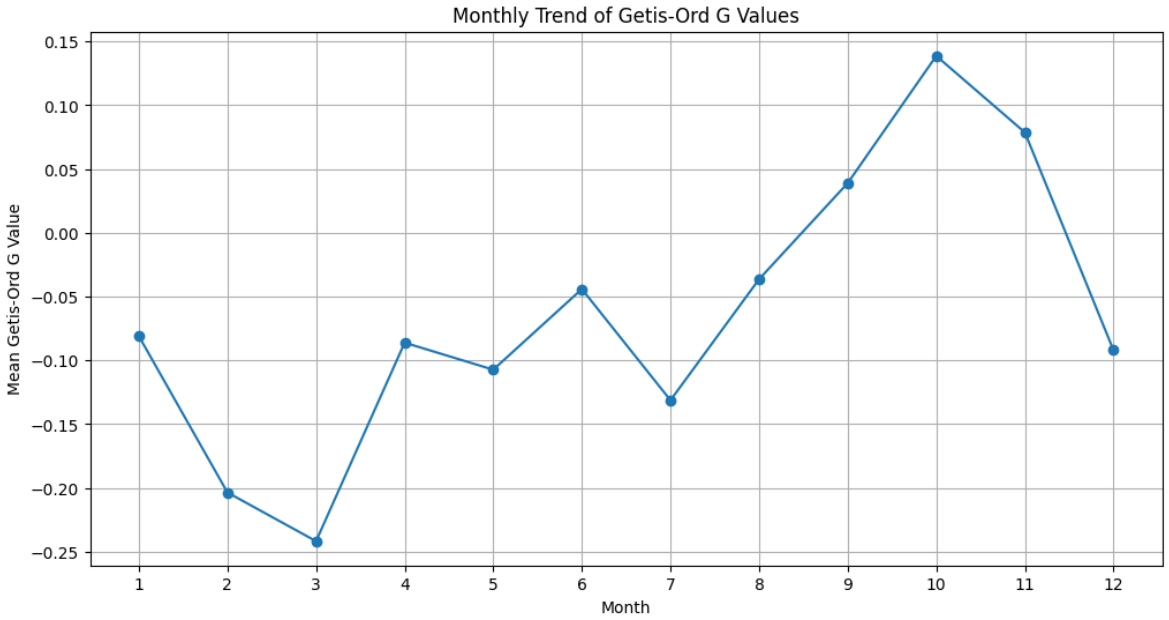

Figure 6. Monthly trend of the G* statistic values.

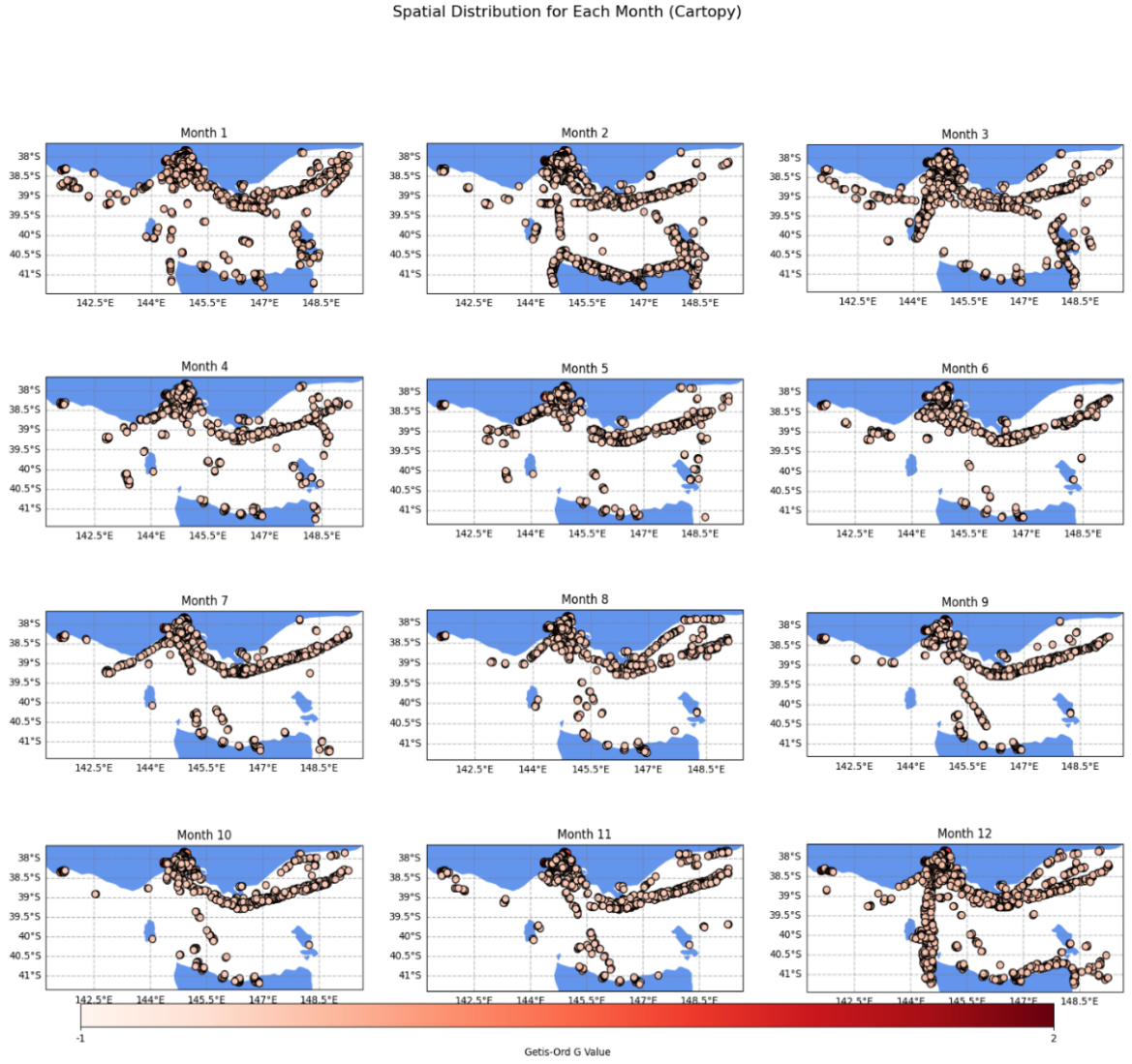

Figure 7. Spatial hotspot distribution for each month of 2022.

4.5. Hourly Trend Analysis and Day-Night Variations

Figure 8 shows the hourly trend and day-night variations of the average Gi* statistics. It can be seen that the values gradually increase during the night hours from midnight to 6 AM, indicating the development of higher risk collision hotspots as the night progresses. At 6 AM, the Gi* statistics peak, indicating the highest collision potential during these early morning hours. After this peak, the values gradually decrease, reflecting a reduction in collision risk. These patterns allow ship officers to better understand the hourly evolution of ship collision risk in the Bass Strait waters and to take proactive measures to reduce ship collision risk. On the other hand, this study divides the day into bi-hourly intervals and visualizes the collision hotspots for each interval in Figure 9. This bi-hourly visualization clearly shows how ship collision risk varies over space and over time. For example, the primary hotspot in the central Bass Strait, visible in the early hours, begins to shift towards Melbourne between 10 AM and 4 PM. This movement reflects regular shipping patterns, with vessels heading towards Melbourne for docking and other activities during this time. After 4 PM, there is an increased collision risk near Burnie port due to increased activity such as cargo operations and fishing returns in the late afternoon and early evening.

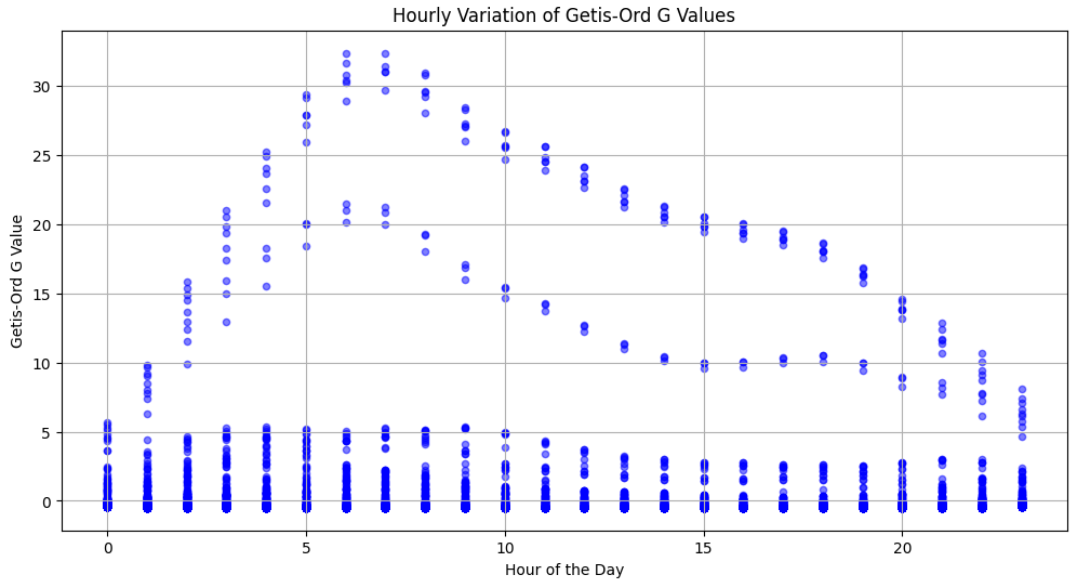

Figure 8. Hourly trend of the G* statistic values.

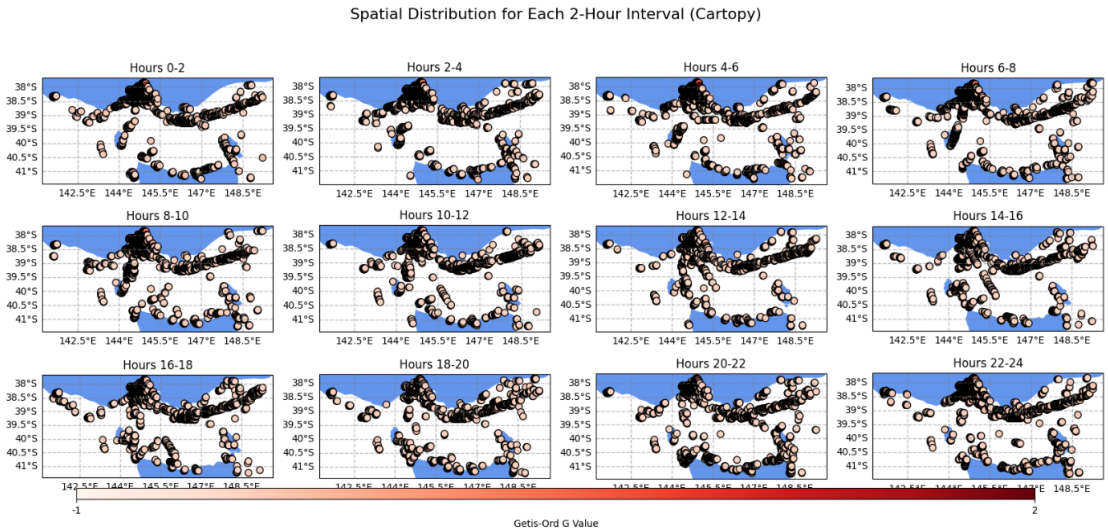

Figure 9. Spatial hotspots per 2 h.

5. Conclusions and Implications

This study provides a comprehensive examination of ship collision risk in the Bass Strait waters using a spatio-temporal hotspot analysis framework, which is enhanced by big data methods and space-time cube technology. The main results are as follows:

First, the analysis shows that most Gi* statistics are clustered around zero, with many being negative. This indicates a lack of significant spatial clustering, highlighting the prevalence of coldspots and only a few areas with extremely high collision risk in the Bass Strait waters monthly. Second, the cluster analysis identifies five high collision risk areas based on spatial and temporal coordinates, as well as the Gi* statistics. These include the blue cluster for the maritime corridor between Sydney and Melbourne, the orange cluster for the Melbourne port precinct, the green cluster for the shipping routes between Melbourne and Tasmania, the red cluster along the northeastern Tasmanian coast, and the purple cluster for the routes between Melbourne and Adelaide. Third, ship collision risk in the Bass Strait waters exhibits clear seasonal and day-night variations, as evidenced by the monthly and hourly trends of the average Gi* statistics. The values are particularly low from December to August, typically remaining negative, but become positive in September, October and November. Additionally, the risk gradually increases during the night hours from midnight to 6 AM, peaking at 6 AM before gradually decreasing throughout the day.

Understanding the spatio-temporal characteristics of ship collision risk provides stakeholders with several valuable insights. First, it enhances stakeholders' abilities to assess ship collision accidents in a timely manner and to respond proactively by rerouting vessels and implementing additional safety measures. Second, it enables stakeholders to allocate resources effectively. For example, port authorities can increase the presence of patrol boats to monitor and direct vessel traffic, and use advanced navigational tools in high-risk areas to help vessels navigate safely by identifying ship collision risk hotspots. Third, it enables stakeholders to develop training modules based on simulation. For example, port personnel can better understand the challenges and patterns of ship collision risk in their operational environment, enabling them to make informed and data-driven decisions in real-time situations. By visualizing ship collision risk hotspots, crew members can better prepare for potential ship collision accidents and develop standard vessel operating procedures and navigational practices. Finally, identifying ship collision risk hotspots enables policymakers to strategically prioritize areas that need immediate attention, ensuring a focused allocation of regulatory resources. The continuous flow of AIS data allows for ongoing evaluations of existing policies to ensure their continued relevance and effectiveness. In summary, the findings of this study can significantly enhance the safety and efficiency of maritime operations by supporting proactive risk management and aiding strategic decision making and policy development.

Despite the valuable implications outlined above, this study can still be extended in different ways. First, this study relies on secondary data from the AMSA and the streamlining of the data may result in the loss of finer details, affecting the overall accuracy of the dataset. It must also be acknowledged that the *CRI* model used has inherent limitations and may introduce bias. Second, the monthly/hourly space-time cube aggregates 10-min CRIs, which limits its ability to capture continuous interaction dynamics or evasive manoeuvres occurring at sub-slice resolution. Near-miss events unfolding between timestamps may therefore be under-represented. Additionally, human and organizational factors (e.g., fatigue, bridge resource management) are not encoded within the *CRI* and, consequently, remain invisible to the Gi* analysis. Third, unlike the VCRO/KF-VCRO approaches that focus on ranking the severity of individual encounters, this study emphasizes the spatial clustering of aggregated encounter risk [9,11]. While the CRI → Gi* framework effectively identifies where and when encounter-risk clusters form, it does not infer causal mechanisms or assess the relative effectiveness of different countermeasures. Such insights can be obtained through Bayesian/BN-TOPSIS models, enabling us to learn causal dependencies and provide decision-oriented rankings [30]. A promising extension is to add a decision layer that encodes causal dependencies among traffic, human, organizational, and environmental factors, updates with new evidence, and uses BN-derived mutual-information weights to rank interventions (e.g., speed management, pilotage, routing, crew training, and early-warning) [30]. Applied to the seasonal and diurnal patterns identified in this study, this transforms static hotspot maps into dynamic, actionable policy tools that bridge risk mapping and decision-making under uncertainty.

Author Contributions

J.Y.: conceptualization, methodology, software, data curation, validation, writing—original draft preparation; W.S.: visualization, investigation, supervision; J.F. and P.B.: writing—reviewing and editing. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data will be made available on request.

Acknowledgments

We appreciate the valuable comments of the anonymous reviewers and Editor, which helped us significantly improve the quality of the research.

Conflicts of Interest

The authors declare no conflict of interest.

Use of AI and AI-assisted Technologies

During the preparation of this work, the authors used DeepL Write to proofread the article. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

1. Wang, H.; Liu, Z.; Wang, X.; et al. An analysis of factors affecting the severity of marine accidents. *Reliab. Eng. Syst. Saf.* **2021**, *210*, 107513.
2. Li, H.; Çelik, C.; Bashir, M.; et al. Incorporation of a global perspective into data-driven analysis of maritime collision accident risk. *Reliab. Eng. Syst. Saf.* **2024**, *249*, 110187.
3. Wu, J.; Thorne-Large, J.; Zhang, P. Safety first: The risk of over-reliance on technology in navigation. *J. Transp. Saf. Secur.* **2022**, *14*, 1220–1246.
4. Li, S.; Meng, Q.; Qu, X. An Overview of Maritime Waterway Quantitative Risk Assessment Models. *Risk Anal.* **2011**, *32*, 496–512.
5. Rong, H.; Teixeira, A.P.; Guedes Soares, C. Spatial correlation analysis of near ship collision hotspots with local maritime traffic characteristics. *Reliab. Eng. Syst. Saf.* **2021**, *209*, 107463.
6. Winkle, T. Safety Benefits of Automated Vehicles: Extended Findings from Accident Research for Development, Validation and Testing. In *Autonomous Driving*; Maurer, M., Gerdes, J., Lenz, B., et al., Eds.; Springer: Berlin/Heidelberg, Germany, 2016; pp. 335–364.
7. Qu, X.; Meng, Q.; Suyi, L. Ship collision risk assessment for the Singapore Strait. *Accid. Anal. Prev.* **2011**, *43*, 2030–2036.
8. Silveira, P.A.M.; Teixeira, A.P.; Soares, C.G. Use of AIS Data to Characterise Marine Traffic Patterns and Ship Collision Risk off the Coast of Portugal. *J. Navig.* **2013**, *66*, 879–898.
9. Zhang, W.; Goerlandt, F.; Montewka, J.; et al. A method for detecting possible near miss ship collisions from AIS data. *Ocean. Eng.* **2015**, *107*, 60–69.
10. Zheng, K.; Chen, Y.; Jiang, Y.; et al. A SVM based ship collision risk assessment algorithm. *Ocean. Eng.* **2020**, *202*, 107062.
11. Liu, Z.; Zhang, B.; Zhang, M.; et al. A quantitative method for the analysis of ship collision risk using AIS data. *Ocean. Eng.* **2023**, *272*, 113906.
12. Moran, P. Notes on Continuous Stochastic henomena. *Biometrika* **1950**, *37*, 17–23.
13. Ord, J.K.; Getis, A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geogr. Anal.* **1995**, *27*, 286–306.
14. Tsou, M.C. Discovering Knowledge from AIS Database for Application in VTS. *J. Navig.* **2010**, *63*, 449–469.
15. Zhang, L.; Meng, Q.; Fang, T.F. 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.
16. Romano, B.; Jiang, Z. Visualizing Traffic Accident Hotspots Based on Spatial-Temporal Network Kernel Density Estimation. In Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Redondo Beach, CA, USA, 7–10 November 2017.
17. Kang, Y.; Cho, N.; Son, S. Spatiotemporal characteristics of elderly population's traffic accidents in Seoul using space-time cube and space-time kernel density estimation. *PLoS ONE* **2018**, *13*, e0196845.
18. Huang, X.; Wen, Y.; Zhang, F.; et al. A review on risk assessment methods for maritime transport. *Ocean. Eng.* **2023**, *279*, 114577.

19. Zhang, J.; Teixeira, Â.P.; Guedes Soares, C.; et al. Quantitative assessment of collision risk influence factors in the Tianjin port. *Saf. Sci.* **2018**, *110*, 363–371.
20. Hänninen, M. Bayesian networks for maritime traffic accident prevention: Benefits and challenges. *Accid. Anal. Prev.* **2014**, *73*, 305–312.
21. Svanberg, M.; Santén, V.; Hörteborn, A.; et al. AIS in maritime research. *Mar. Policy* **2019**, *106*, 103520.
22. Yang, D.; Wu, L.; Wang, S.; et al. How big data enriches maritime research—A critical review of Automatic Identification System (AIS) data applications. *Transp. Rev.* **2019**, *39*, 755–773.
23. Tritsarolis, A.; Chondrodima, E.; Pelekis, N.; et al. Vessel Collision Risk Assessment using AIS Data: A Machine Learning Approach. In Proceedings of the 23rd IEEE International Conference on Mobile Data Management (MDM), Paphos, Cyprus, 6–9 June 2022; pp. 425–430.
24. AMSA (Australian Maritime Safety Authority). Digital Data. Available online: https://www.operations.amsa.gov.au/Spatial/DataServices/DigitalData (accessed on 5 July 2023).
25. Park, J.; Jeong, J.S. An Estimation of Ship Collision Risk Based on Relevance Vector Machine. *J. Mar. Sci. Eng.* **2021**, *9*, 538.
26. Dong, W.; Zhang, P.; Li, J. Safety First—A Critical Examination of the Lights and Shapes in COLREGs. *J. Mar. Sci. Eng.* **2023**, *11*, 1508.
27. Nakaya, T.; Yano, K. Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics. *Trans. GIS* **2010**, *14*, 223–239.
28. Zhang, M.; Montewka, J.; Manderbacka, T.; et al. A Big Data Analytics Method for the Evaluation of Ship–Ship Collision Risk reflecting Hydrometeorological Conditions. *Reliab. Eng. Syst. Saf.* **2021**, *213*, 107674.
29. Sail-world. Glorious forecast for start of 50th anniversary of Melbourne to Hobart Yacht Race. Available online: https://www.sail-world.com/news/257123/Glorious-forecast-for-start-of-50th-M2H-Yacht-Race (accessed on 25 June 2023).
30. Fan, H.; Lu, J.; Chang, Z.; et al. A Bayesian network based—TOPSIS framework to dynamically control the risk of maritime piracy. *Marit. Policy Manag.* **2024**, *51*, 1582–1601.