Evolution trends and influencing factors analysis for the severity and pollution of maritime accidents in Arctic waters from multi-source data

Downloaded from: https://research.chalmers.se, 2025-09-25 13:42 UTC

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

Fu, S., Cui, M., Wu, N. et al (2026). Evolution trends and influencing factors analysis for the severity and pollution of maritime
accidents in Arctic waters from multi-source data. Reliability Engineering and System Safety, 266.
http://dx.doi.org/10.1016/j.ress.2025.111644

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

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

Reliability Engineering and System Safety 266 (2026) 111644

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Evolution trends and influencing factors analysis for the severity and pollution of maritime accidents in Arctic waters from multi-source data

Shanshan Fu [a,c], Mengfei Cui [a], Ningji Wu [a], Mingyang Zhang [b,*], Xiao Lang [c], Wengang Mao [c]

[a] College of Transport & Communications, Shanghai Maritime University, Shanghai 201306, China
[b] MOE Key Laboratory of Marine Intelligent Equipment and System, Shanghai Jiao Tong University, Shanghai, 200240, China
[c] Department of Mechanics and Maritime Sciences, Chalmers University of Technology, Gothenburg 41296, Sweden

ARTICLE INFO

Keywords:
Maritime accident
Arctic shipping
Evolution trends analysis
Bivariate probit model

ABSTRACT

As Arctic maritime activity has increased, Arctic shipping risk management has attracted extensive attention. Given the limited availability of rescue equipment in Arctic waters, maritime accidents in ice-covered regions tend to result in more severe consequences and a greater environmental impact. This paper proposes a data-driven framework for analyzing evolution trends and identifying the influencing factors of maritime accidents in Arctic waters, in terms of both their severity (casualties and property loss) and pollution (environmental impact). First, a data preparation approach is proposed to integrate maritime accident data and associated hydrometeorological data from 2004 to 2023 for Arctic waters covered under the Arctic Search and Rescue Agreement. Second, the evolutionary trends of maritime accidents in Arctic waters are analyzed in terms of their severity and environmental pollution. Third, a bivariate probit model is proposed to explore the factors affecting the severity and environmental pollution of maritime accidents in Arctic waters. Finally, a marginal effect analysis is conducted to quantify the impacts of these factors on the severity and environmental pollution. The results indicate that flag state characteristics significantly influence both the severity and the pollution of accidents. Additionally, factors such as machinery damage, wrecked, allision, hull damage, fire/explosion, strong winds, and sea ice contribute positively to the severity of accidents, while these factors negatively influence pollution accidents to varying degrees.

1. Introduction

In recent years, maritime activities in Arctic waters have increased due to the melting of sea ice [1–3]. Arctic shipping risk management faces challenges including harsh environmental conditions, inadequate infrastructure, and ecological vulnerability [4,5]. More specifically, the serious casualties and pollution caused by accidents in ice-covered waters are exacerbated by the limited availability of search and rescue resources in Arctic regions. Therefore, a comprehensive analysis of maritime accidents in Arctic waters is essential for developing targeted risk control options (RCOs) and enhancing Arctic shipping risk management.

Previous research on Arctic shipping risk management has primarily focused on estimating the probability of maritime accidents caused by the individual or combined effects of sea ice conditions (e.g., ice thickness and concentration), weather conditions (e.g., wind, visibility), and ship operations (e.g., ship speed, engine power) [6–9]. The various types of maritime accidents include ships being beset in ice [10–12], ship-ice collisions [13], grounding [14,15], and ship-ship collisions during convoy operations [16]. Fu et al. [17] and Xu et al. [18] developed probabilistic Bayesian network (BN) models for the prediction of ship besetting in ice using real Arctic shipping data from the Northeast Passage (NEP). Khan et al. [13,19] predicted the probability of ship-ice collisions by considering weather and sea ice conditions. Fu et al. [15] proposed a quantitative risk assessment framework for grounding accidents using AcciMap-BN. Liu et al. [20] and Zhu et al. [21] investigated risk influencing factors and estimated the navigational risk of convey operations using multi-source data. Fu et al. [22] developed an object-oriented BN model for the quantitative risk assessment of multiple navigational accidents in ice-covered Arctic waters. Among these studies, BNs have been the most popular modeling technique [23] due to their flexibility in integrating statistical data and experimental

* Corresponding author at: Shanghai, 200240, China.
E-mail address: mingyang.zhang@sjtu.edu.cn (M. Zhang).

https://doi.org/10.1016/j.ress.2025.111644
Received 24 April 2025; Received in revised form 12 August 2025; Accepted 31 August 2025
Available online 1 September 2025
0951-8320/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

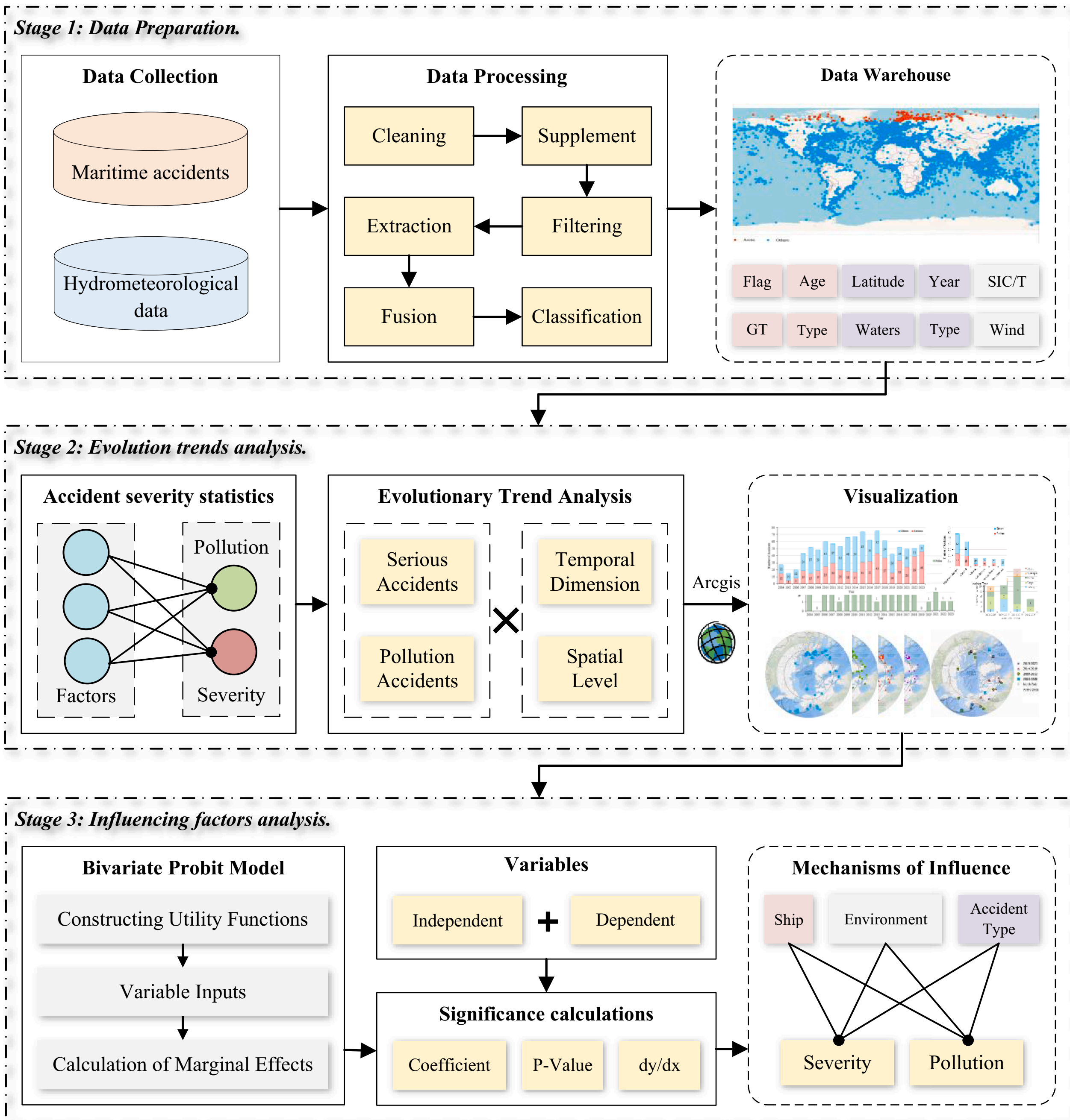


Fig. 1. The framework for maritime accident analysis in Arctic waters.

judgment, particularly in the early stage of Arctic shipping risk management. With the increasing availability of hydrometeorological and maritime accident data, statistical models such as logistic regression are being used more frequently to identify key factors influencing maritime accidents [20]. It proposes an intelligent fleet kinematic model that integrates a Traffic State Network (TFSN) with an improved PID controller, for the first time comprehensively incorporating mechanical delay, communication delay, ship transmission delay, and ice resistance into ship safety state classification and speed control, enabling dynamic simulation of safe speeds and distances for multiple ships under ice-breaking coordination and generating reliable formation maneuver commands [24]. However, more attention should be paid to the evolution trends of maritime accidents under Arctic climate change and the specific effects of various technical and environmental factors on casualties and the environmental impact of maritime accidents.

In general, maritime accident analysis research has increasingly focused on temporal and spatial characteristics [25–27] and the identification of key factors influencing the severity of maritime accidents [28,29]. For example, Li et al. [30] explored the dynamic evolution of maritime accidents and associated influencing factors using the International Maritime Organization (IMO) Global Integrated Shipping Information System (GISIS) database. Zhang et al. [27,31] discussed the spatial distribution of global maritime accidents and piracy incidents using text mining and geospatial techniques. Sui et al. [32] conducted time series analysis of maritime accidents considering their spatial distribution. Çakır et al. [33] investigated the severity of maritime accidents using association rule mining; Zhou et al. [34] studied the spatiotemporal patterns of maritime accidents. Li et al. [35] investigated spatial heterogeneity using a geographically weighted regression model. Lau et al. [36] analyzed the pollution risk of maritime accidents using

Table 1
Data sources.

Attribute	Dataset (columns)	Source (link)
Ship	Ship ID (IMO and MMSI numbers)	
	Ship type	• Lloyd's List Intelligence (https://www.lloydslistintelligence.com)
	Flag	• Clarksons (https://www.clarksons.com)
	Age	• Equasis (https://www.equasis.org)
	Gross tonnage (GT)	
Accident	Timestamp (year-month-day-hour)	• Lloyd's List Intelligence (https://www.lloydslistintelligence.com)
	Location (Longitude and latitude)	• IMO GISIS (https://gisis.imo.org/public/default.aspx)
	Accident type	
	Severity (Casualties and property loss)	
	Pollution	
Environment	Sea ice concentration (SIC)	• University of Bremen (https://data.seaice.uni-bremen.de/databrowser)
	Sea ice thickness (SIT)	• ERA5 (https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5)
	Wind	

port state control inspection data. Feng et al. [28] applied machine learning techniques to predict accident severity. Cao et al. [37] integrated association rule mining and complex network analysis to explore the risk factors influencing maritime accidents. Chen et al. [38] and Wan et al. [39] analyzed pollution accidents specific to oil tankers and container ships, respectively. These studies have primarily relied on maritime-related databases, utilizing regression models [29], network models [40,41], and geospatial techniques [31] to comprehensively analyze the spatiotemporal characteristics and influencing factors of ship and accident-related maritime accidents. Liu et al. [42] proposed a deep ensemble learning model based on neural oblivious decision trees (NODE), combined with focal loss and MC dropout–based uncertainty quantification, achieving 97% accuracy, 95% precision, and 93% recall on highly imbalanced navigation-mode data, significantly outperforming baseline models such as random forest and gradient boosting, and enabling spatiotemporally scalable navigation-mode prediction maps to support intelligent decision-making and optimize icebreaker resource allocation. However, existing studies have primarily focused on identifying factors that influence severe accidents or pollution accidents, and limited studies have quantified the impact of accidents or conducted comprehensive analyses of factors affecting casualties, property damage, and environmental impact. Moreover, hydrometeorological data related to maritime accidents have been insufficiently considered in maritime accident analysis.

To address these gaps, it is essential to comprehensively analyze the dynamic evolution of such accidents and identify the factors that influence both the severity (casualties and property loss) and pollution (environmental impact) of maritime accidents in Arctic waters. Therefore, this paper proposes an analytical framework that encompasses data preparation, evolution trends analysis, and influencing factors analysis of maritime accidents in Arctic waters by integrating maritime accident data with hydrometeorological data. A bivariate probit model is developed to identify key factors affecting the severity and pollution of maritime accidents, considering the coupling relationship between these two consequences. Moreover, the specific influences of these factors on the severity and pollution of accidents are quantified using marginal effect analysis. The proposed framework and methods offer a novel perspective on Arctic shipping risk management, highlighting the most critical influencing factors for effective risk mitigation.

The rest of the study is organized as follows: Section 2 describes the data sources and methods used to characterize the accidents, Section 3 describes the preprocessing of accident, sea ice, and wind data, Section 4 analyzes the evolutionary characteristics and trends of accidents in Arctic waters from a temporal and spatial perspective, Section 5 explores the intrinsic mechanisms behind accident impacts using a bivariate probit model, Section 6 further calculates marginal effects to quantify the impact of key factors on accidents, and finally, Section 7 summarizes the results and presents potential directions for future research.

2. Methodology

2.1. Framework

This paper proposes a data-driven framework for analyzing evolution trends and identifying the factors that impact both the severity (casualties and property loss) and pollution (environmental impacts) of maritime accidents in Arctic waters by integrating maritime accident data with hydrometeorological data. This framework includes data preparation, evolution trends analysis, and influencing factors analysis, as shown in Fig. 1.

2.2. Stage 1: Data preparation

In order to obtain a high-quality dataset for maritime accident analysis, several data sources were used to collect, process, and supplement maritime-accident-related datasets: Lloyd's List Intelligence (https://www.lloydslistintelligence.com/), IMO GISIS (https://gisis.imo.org/public/default.aspx), the University of Bremen Sea Ice dataset (https://data.seaice.uni-bremen.de/databrowser/), Clarksons (https://www.clarksons.com), and the European Center for Medium-range Weather Forecasts (ECMWF) ERA5 atmospheric reanalysis dataset (https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5). The final dataset comprised three categories of attributes: ship information (e.g., type, flag, age, GT), maritime accident details (timestamp, location, type, severity, pollution), and environmental conditions (e.g., SIC, SIT, wind), as shown in Table 1.

The original maritime accident dataset was collected from Lloyd's List Intelligence. The data preparation process is shown in Fig. 2 and comprises the following six steps:

Step 1: Data Cleaning. Using the collected global maritime accident dataset from Lloyd's List Intelligence, each accident record was individually checked for accuracy. If an accident record contained missing, incorrect, or unreasonable information (e.g., a missing timestamp or location, or inconsistent severity classification, such as an accident labeled as a total loss yet not serious), that specific accident record was removed from the dataset. It was ensured that all maritime accident entries had location information (latitude and longitude). Moreover, the IMO GISIS database was used to verify the accuracy of the accident information.

Step 2: Data Filtering. The maritime accident dataset was imported into the ArcGIS Pro software [43–49] to extract accidents in Arctic waters corresponding to regions defined in the Arctic Search and Rescue Agreement [50], which is represented as an irregular polygon.

Step 3: Data supplementation. Some ship information (e.g., age, GT, and ship type) in the maritime accident dataset was omitted in the Lloyd's List Intelligence. These data were added using the Ship ID (IMO and MMSI numbers) to retrieve data from the Equasis and Clarksons datasets.

Step 4: Data Extraction. Based on the timestamp (year-month-day-hour) and location (longitude and latitude) of the maritime accidents, the corresponding environmental conditions (e.g., SIC, SIT, and wind) were extracted from the University of Bremen Sea Ice and ERA5 datasets.

Step 5: Data discretization. Some continuous variables were discretized to capture evolutionary trends and support subsequent discrete choice modeling.

Step 6: Data Storage. The processed data were stored, and a new warehouse was created to analyze subsequent evolution trends and

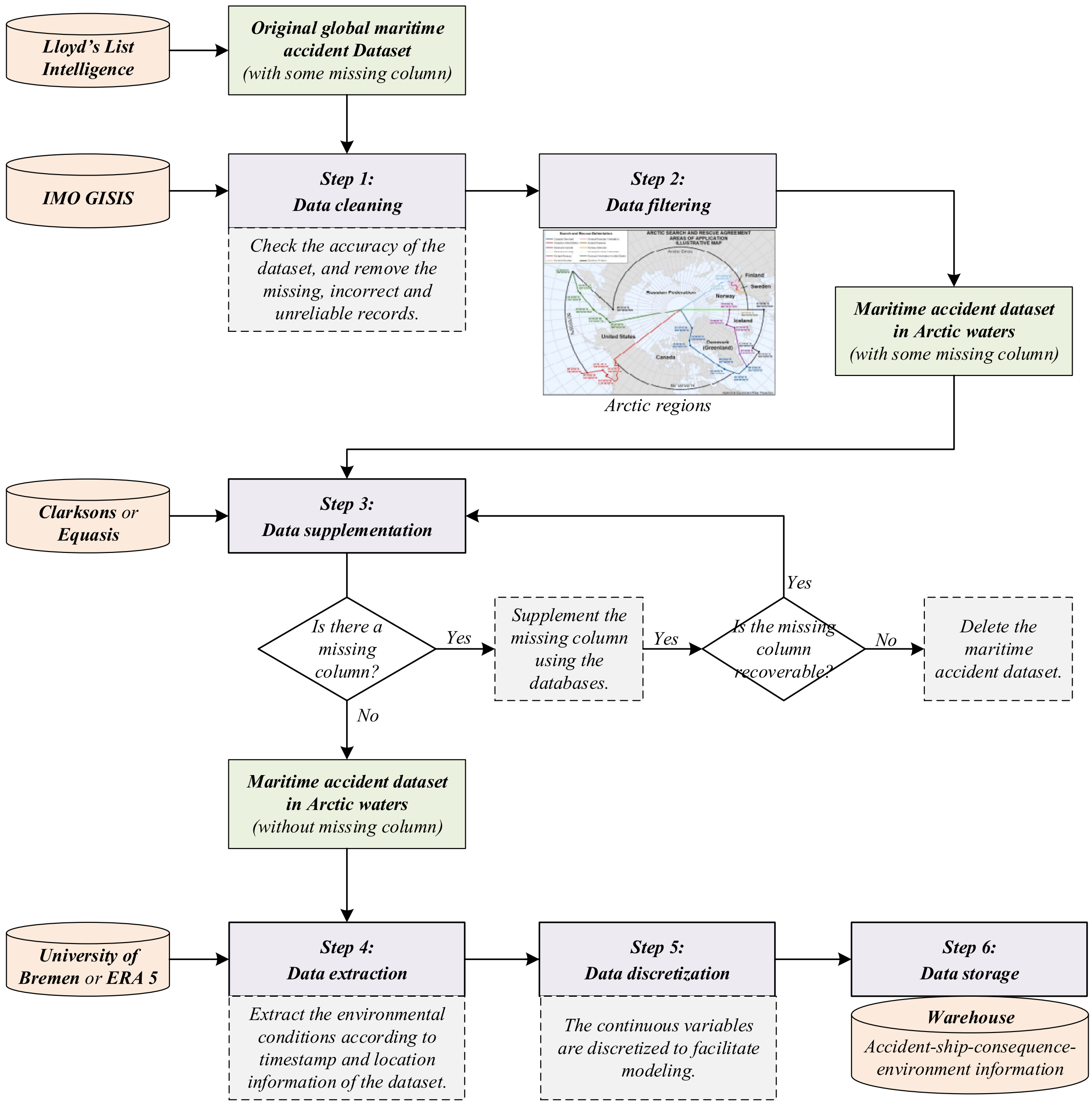


Fig. 2. Data preparation process.

influencing factors. The final data warehouse for the maritime accident dataset in Arctic waters includes the following columns: multimodal state (ship type, flag, age, accident year, accident waters, accident type, wind) and binary (severity, pollution, SIC, SIT).

2.3. Stage 2: Evolution trends analysis

The evolutionary trends of the accident severity and pollution were analyzed by dividing the dataset into four intervals, every five years on average, to explore the evolutionary features of maritime accidents, their associated influencing factors, and their geospatial aspects in different periods.

- Trends of annual evolution. Using the processed Arctic accident data, annual statistical analyses were separately performed for the number of severity and pollution accidents, and trend lines were fitted using second-order polynomials.
- Temporal evolution of ship-related factors. Statistical analysis of the types of ships, such as fishing or cargo ships, that cause severe accidents or accidents with significant pollution was conducted in four annual intervals to explore the trends in the frequency of accidents on each type of ship.
- Temporal evolution of accident-related factors. Statistical analysis of the types of initial accidents within the four annual intervals, such as mechanical failures or fires/explosions, was conducted to explore the trends in the frequency of various types of accidents.
- Spatiotemporal evolution of accidents. Using Geographic Information System (GIS) software such as ArcGIS Pro, severity and pollution accidents occurring in the four annual intervals were visualized with their latitude and longitude. The spatial evolutionary trend of accidents was also analyzed in terms of the density of accident sites in the watershed.

2.4. Stage 3: Influencing factors analysis

The discrete choice model [51] is a widely used econometric framework for data analysis, particularly in traffic accident research, where it facilitates the modeling of individual decision-making processes. Commonly applied models include the multinomial logit, probit, and mixed logit, each of which offers distinct advantages in capturing unobserved heterogeneity, correlation structures, and choice

Table 2
Sources and resolution of the dataset.

Parameter	Source	Section	Spatial Resolution	Temporal Resolution	Reference
SIC	University of Bremen	[0, 100] %	12.5 × 12.5 km	Daily	[53]
SIT	University of Bremen	[0, 51] cm	6.25 × 6.25 km	Daily	[54]
Wind	ERA5	[0, +∞) m/s	0.5 × 0.5 degree	Hourly	[55]

probabilities in accident-related studies. The bivariate probit model can couple the calculation of the severity and pollution of accidents, which has significant practical significance.

A bivariate probit model [52] is a joint model of the outcomes of two binary variables. Bivariate probit models are used when there are two outcome variables in a model. It is assumed that there is a correlation between the stochastic perturbation terms of the system of equations and that the equations in the model must be estimated simultaneously. If the outcomes of the two binary variables are uncorrelated, we can apply two separate probit models. If the outcomes of the two binary variables are correlated and the use of a probit model leads to biased estimates and affects the conclusions, a bivariate probit model can be used.

The object of this study is the correlation between severity accidents and pollution accidents. Therefore, a bivariate probit model can be applied. The utility function of the model is defined as:

$$\begin{cases} Y_1^* = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \varepsilon_1 \\ Y_2^* = \boldsymbol{X}_2\boldsymbol{\beta}_2 + \varepsilon_2 \end{cases}, \tag{1}$$

where Y_1^* and Y_2^* denote the dependent variables of the severity and pollution, respectively. X_1 and X_2 are the explanatory variables corresponding to Y_1^* and Y_2^*, which may or may not be the same. β_1 and β_2 are the parameter vectors, and the perturbation terms ε_1 and ε_2 are subject to a joint standard normal distribution, with an expectation of 0, a variance of 1, and a correlation coefficient of ρ used to measure the correlation between the two equations.

$$(\varepsilon_1, \varepsilon_2) \sim N(0,\ 0,\ 1,\ 1,\ \rho) \tag{2}$$

If the latent variables Y_1^* and Y_2^* are greater than 0, the dependent variables Y_1 and Y_2 are observed. The observable variables Y_1 and Y_2 are defined as follows:

$$Y_1 = \begin{cases} 1, Y_1^* > 0 \text{ (severity accident)} \\ 0,\ \textit{others} \text{ (not severity accident)} \end{cases} \tag{3}$$

$$Y_2 = \begin{cases} 1,\ Y_2^* > 0 \text{ (pollution accident)} \\ 0,\ \textit{others} \text{ (notpollution accident)} \end{cases} \tag{4}$$

When the explanatory variables of the two equations of Eq. (1) are identical, i.e., $X_1 = X_2$, the result is a bivariate probit model with identical variables. Conversely, when the explanatory variables of the two equations are not identical, i.e., $X_1 \neq X_2$, the model is called a seemingly uncorrelated bivariate probit model because the only link between the two equations in the model is the correlation of the perturbation terms. If $\rho = 0$, the model is equivalent to two separate probit models. When $\rho \neq 0$, the probabilities of $(Y_1,\ Y_2)$ can be written down and then estimated by maximum likelihood. Since the error terms ε_1 and ε_2 obey a joint normal distribution, the joint probability is:

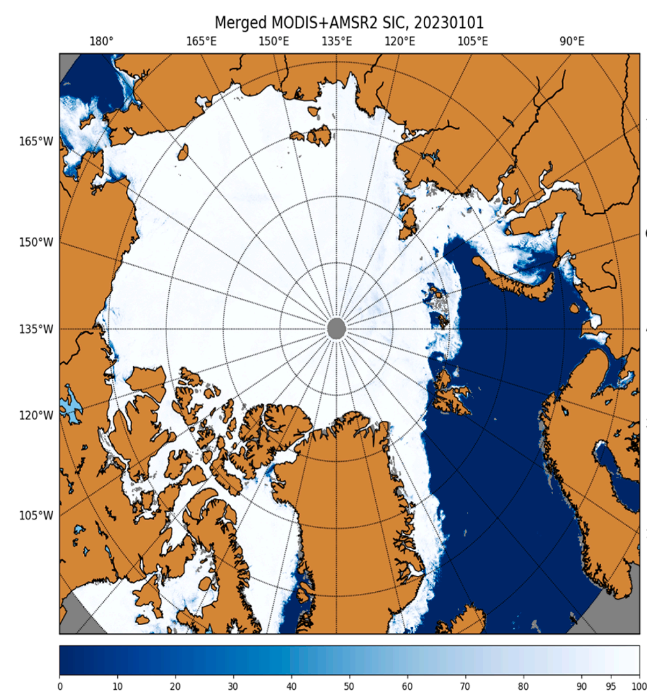
$$P(Y_1 = y_1,\ Y_2 = y_2 | X_1,\ X_2) = \Phi(y_1 X_1 \beta_1,\ y_2 X_2 \beta_2,\ \rho) \tag{5}$$

where y_1 and y_2 take values of 0 or 1.

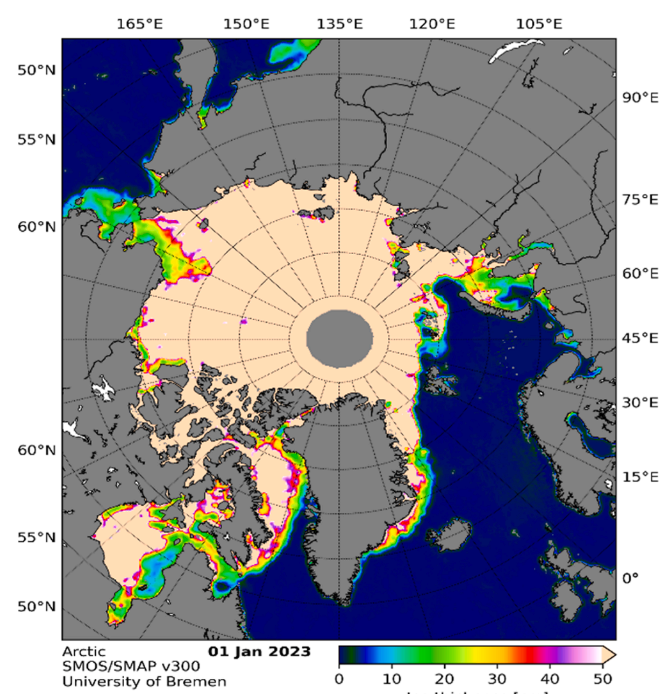
For the joint probability, the marginal effect of each influencing factor is calculated as follows:

$$\frac{\partial P(Y_1 = 1,\ Y_2 = 1 | X)}{\partial X_k} = \Phi_2(X_1\beta_1,\ X_2\beta_2;\ \rho)\left(\beta_{k,1} + \beta_{k,2}\right) \tag{6}$$

In this equation, Φ_2 is the probability density function of the

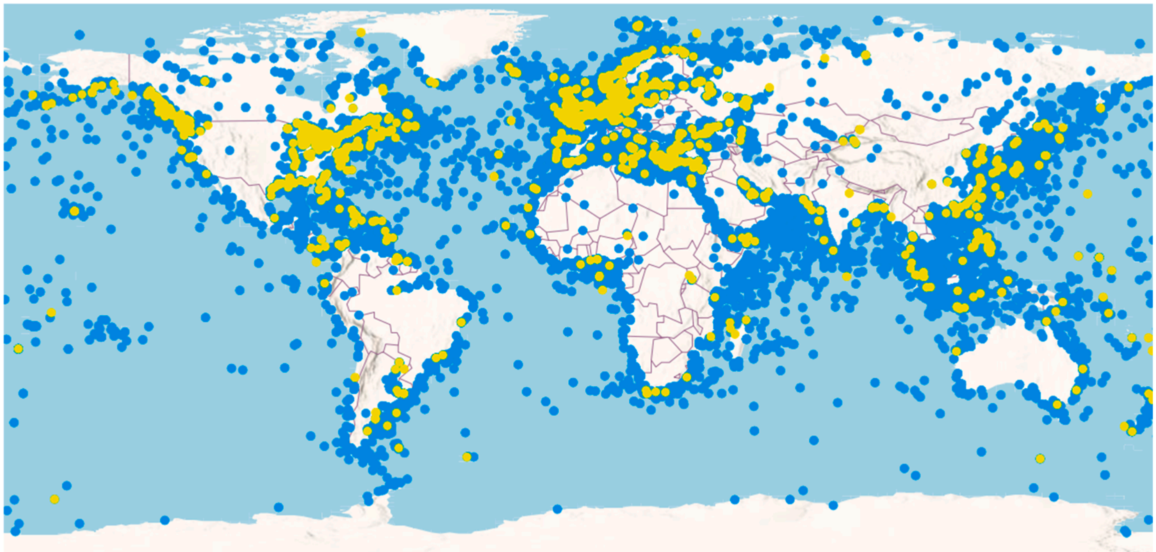


(a) SIC

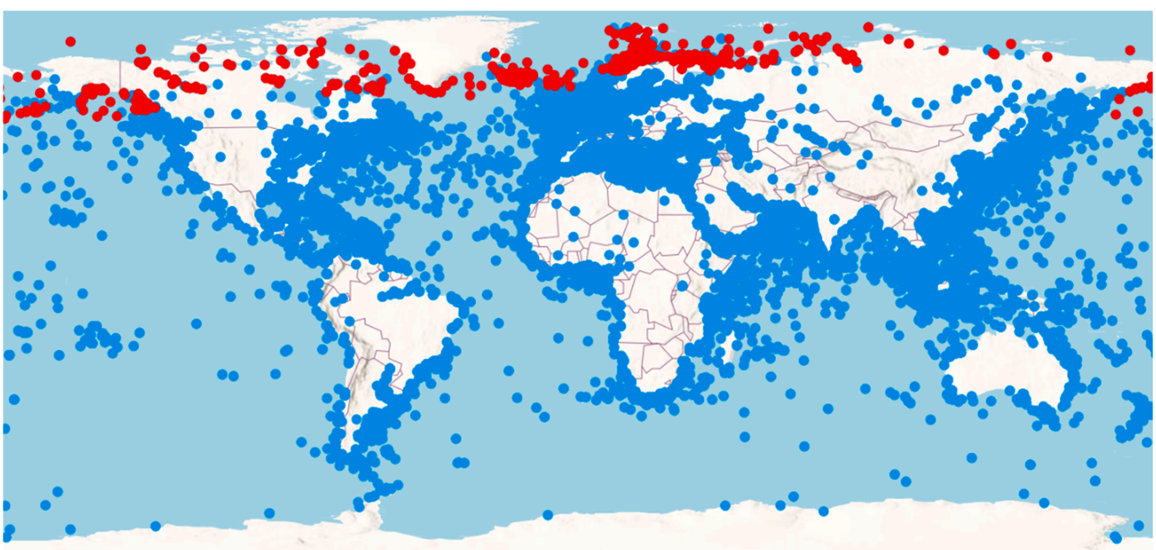


(b) SIT

Fig. 3. An example of SIC and SIT maps from the Bremen dataset (2023-01-01).



(a) The original global maritime accident data from Lloyd's List Intelligence from 2004 to 2023 (56,288 items).



(b) The maritime accidents in Arctic waters for evolution analysis (1,028 items in red).

Fig. 4. Data preparation process.

bivariate normal distribution. $\beta_{k,1}$ and $\beta_{k,2}$ are the regression coefficients of X_k in the two equations, respectively.

3. Data preparation

3.1. Data collection

We collected original maritime accident data from Lloyd's List Intelligence for the period 2004-2023, covering a total of 56,288 accidents. The dataset included ship information (ship IMO, MMSI number, ship name, age, gross tonnage (GT), ship type, built by, built at, class, P & I club, commercial operator, and other information about the ship itself and its ownership) and accident information (including timestamp, accident latitude/longitude, accident type, consequence, waters, and other accident-related information).

The detailed sources and resolution information of the environmental data are listed in Table 2. The environmental data were downloaded as NetCDF files from the University of Bremen Sea Ice and ERA5 datasets, with the Northern Earth domain selected according to accident dates. SIC and SIT data were extracted from the sea ice datasets of the University of Bremen, which are based on satellite observations from MODIS and AMSR2 sensors and provide high-quality data on sea ice status and extent. Examples of SIC and SIT maps from the Bremen dataset are shown in Fig. 3. Wind data, which record the wind speed at a height of 10 meters above the Earth's surface, were extracted from the ERA5 dataset. The SIC data range was 0-100%, and the SIT data range was 1-51 cm (sea ice with a thickness greater than 51 cm was recorded as 51 cm). The observational unit for wind data was meters per second.

3.2. Data processing

Given the large scale of original maritime accident data from 2004 to 2023 (totaling 56,288 items), there was a risk of problems such as missing data, recording errors, or systematic errors in the data collection and recording process. To ensure the accuracy and reliability of the data analysis, this study necessitated the cleaning and processing of the

Table 3
Data classification.

Factors		Classifications	Reference
Ship	Flag	1: Norway, Iceland, Denmark, Finland. 2: Russia. 3: Canada and U.S.A. 4: Others (non-Arctic countries).	[56,57]
	GT	1: (0, 3000) tons. 2: [3000, 10000) tons. 3: [10000, 20000) tons. 4: [20000, $+\infty$) tons.	
	Age	1: (0, 10] years. 2: (10, 20] years. 3: (20, 30] years. 4: (30, $+\infty$) years.	
	Type	1: Fishing. 2: Passenger. 3: Cargo ship. 4: Tanker ship. 5: Others.	
Accident	Year	1: 2004~2008. 2: 2009~2013. 3: 2014~2018. 4: 2019~2023.	
	Waters	1: Iceland and northern Norway. 2: Russia, Arctic, and Bering Sea. 3: Canadian Arctic and Alaska.	
	Latitude	1: Low (Below 66°34' N in Arctic waters). 2: Medium (66°34' ~ 70° N). 3: High (Above 70° N).	
	Type	1: Machinery damage. 2: Wrecked. 3: Allision. 4: Hull damage. 5: Fire/ Explosion. 6: Miscellaneous.	[56,58, 59]
Environment	SIC	1: (0, 0.1) %. 2: [0.1, 1) %.	
	SIT	1: below 10 cm. 2: above 10 cm.	
	Wind	1: (0, 8.0) m/s. 2: [8.0, 13.9) m/s. 3: [13.9, $+\infty$) m/s.	
Consequence	Severity	1: Severity accident (casualties and property loss). 2: Other accident.	[56]
	Pollution	1: Pollution accident (environmental impacts). 2: Other accident.	

Table 4
Ship type descriptions.

Ship factors	Description	Examples
Fishing	Used for fishing activity	Fishing, reefer, fish factory, fish carrier, etc.
Passenger	Ships that carry paying passengers	Passenger, passenger ro/ro, ferry, research, etc.
Cargo ship	Carrying dry bulk, general cargo, containers	General cargo, bulk carrier, fully cellular containership, etc.
Tanker ship	Carrying bulk liquids or chemicals	Tank barge, chemical tanker, combined chemical and oil tanker, liquefied natural gas carrier, etc.
Others	Engineering or icebreaking	Drill ship, tug, icebreaker, drill platform, etc.

Table 5
Accident type descriptions.

Accident factors	Description
Machinery damage	Failure of or damage to the ship's mechanical equipment or power system, resulting in the ship being unable to function or navigate normally.
Wrecked	Loss of the ship's ability to navigate normally as a result of the ship's deviation from its course, causing the ship's bottom to come into contact with the seabed or land.
Allision	Includes collision (a ship striking another ship) and contact (accidental contact between a ship and a non-moving object that is not a ship, such as a pier or bridge).
Fire/Explosion	Fire or explosion caused by an ignition source on board a ship.
Hull damage	Scratches, indentations on the surface of the hull (shell), or rupture of the hull structure.
Miscellaneous	Other unforeseen special maritime accidents that are not readily classifiable.

Table 6
The final data set description.

Factors	Number	Prop. %	Factors	Number	Prop. %
All data	1028	100%	***Severity consequence***		
Ship Flag			Severity	536	52.14%
Norway, etc.	319	31.03%	Others	492	47.86%
Russia	260	25.29%	***Pollution consequence***		
Canada and U.S. A.	183	17.80%	Pollution	45	4.38%
Others	266	25.88%	Others	983	95.62%
Ship GT			***Latitude***		
(0, 3000) tons	560	54.47%	Low	384	37.35%
[3000, 10000) tons	259	25.19%	Medium	441	42.90%
[1000,20000) tons	101	9.82%	High	203	19.75%
[20000, $+\infty$) tons	108	10.51%	***Accident type***		
Ship Age			Machinery damage	431	41.93%
(0,10] years	228	22.18%	Wrecked	185	18.00%
(10,20] years	201	19.55%	Allision	125	12.16%
(20,30] years	270	26.26%	Fire/Explosion	97	9.44%
(30, $+\infty$) years	329	32.00%	Hull damage	52	5.06%
Ship type			Miscellaneous	138	13.42%
Fishing	339	32.98%	***SIC***		
Passenger	290	28.21%	(0, 0.1) %	950	92.41%
Cargo ship	226	21.98%	[0.1, 1) %	78	7.59%
Tanker ship	91	8.85%	***SIT***		
Others	82	7.98%	Below 10 cm	993	96.60%
Year			Above 10 cm	35	3.40%
2004-2008	157	15.27%	***Wind***		
2009-2013	284	27.63%	(0, 8.0) m/s	479	46.59%
2014-2018	339	32.98%	[8.0, 13.9) m/s	383	37.26%
2019-2023	248	24.12%	[13.9, $+\infty$) m/s	166	16.15%

original global maritime accident data (as described in Section 2.2). Initially, a systematic data preprocessing of the original accident data was conducted, eliminating records of negligible reference value (such as those with missing locations or excessively old ships), leaving 53,122 entries. Subsequently, Arctic accidents were filtered from the global maritime accidents, amounting to 1,273 entries. Finally, the original Arctic accident data underwent further supplementation and cleansing, culminating in a dataset of 1,028 ship accidents in the Arctic waters from 2004 to 2023. This dataset was devoid of missing strings but did not include environmental data. The data preparation process is summarized in Fig. 4.

Based on the date and latitude/longitude of the 1,028 Arctic ship accidents, the SIC, SIT, and wind data were extracted from the NetCDF files and fused with the accident dataset according to the time and latitude/longitude. The complete 1,028-entry dataset was finally obtained without any missing strings and with environmental data.

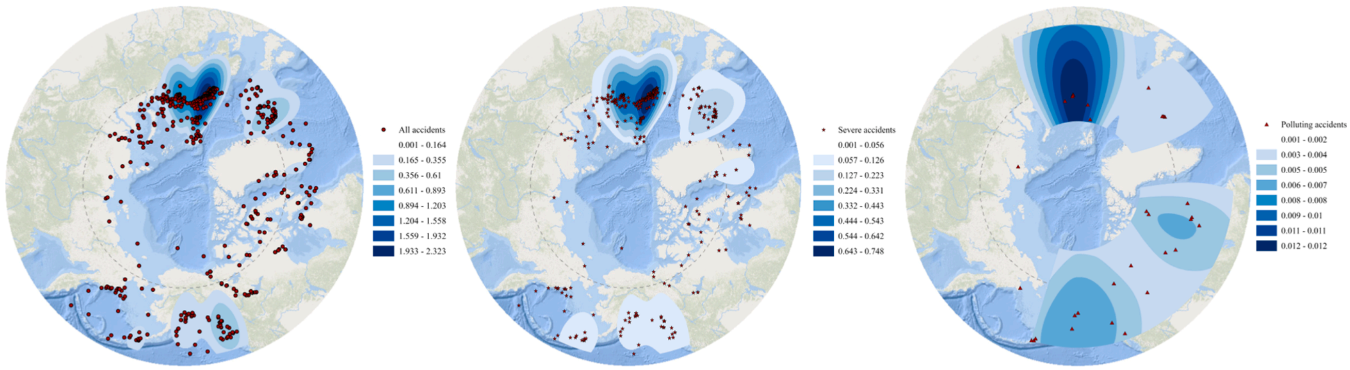
For further analysis, the complete dataset was further categorized by applying the methods of equipartition, literature references, regulations and standards. This classification is shown in Table 3.

Due to the large number of ship types, Table 4 provides a more detailed description of the classification of ship types and summarizes the subcategories contained within the ship types. Accidents are classified into five types using the initial accident type as a reference, and descriptions and explanations of specific accidents are listed in Table 5.

3.3. Data storage

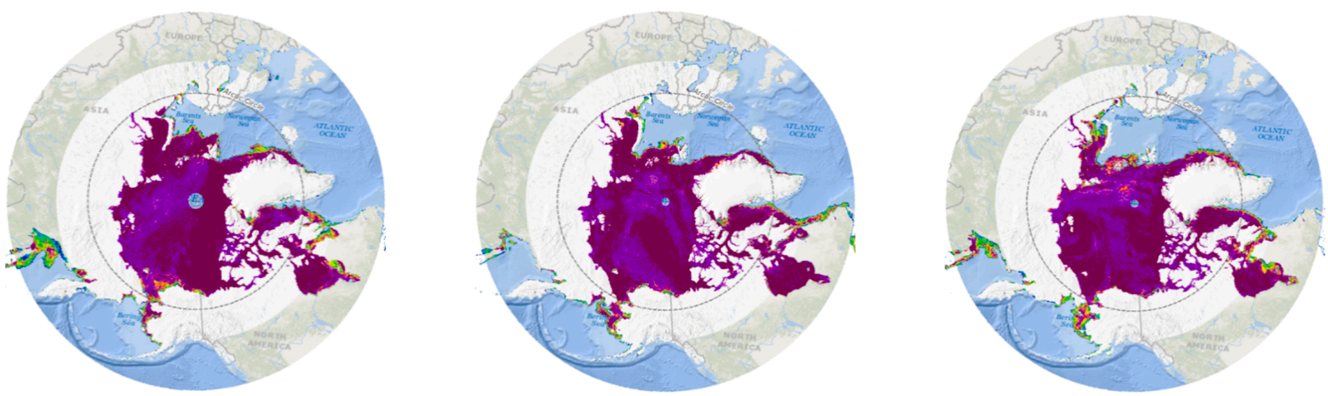
This study considers maritime accident and environmental data for the period 2004-2023, and a multidimensional dataset containing 1,028 complete records was constructed through systematic data cleaning, matching, and fusion processing. The dataset was standardized and classified to provide a reliable data basis for subsequent research on accident evolution trend analysis and impact mechanism analysis. The structural characteristics and frequency distribution of the dataset are shown in Table 6.

The dataset was imported into ArcGIS Pro for visualization of the distribution of accidents in the Arctic region, with the detailed results

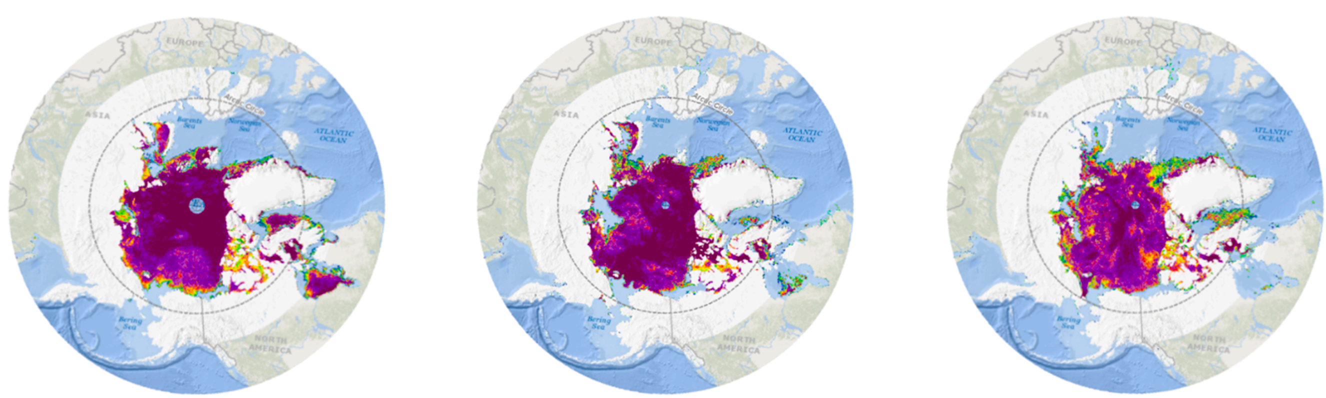


(a) All accidents. (b) Severe accidents. (c) Polluting accidents.

Fig. 5. The spatial distribution of ship accidents in Arctic waters from 2004 to 2023.



(a) 2004-Jan-01. (b) 2014-Jan-01. (c) 2023-Jan-01.



(d) 2004-July-01. (e) 2014-July-01. (f) 2023-July-01.

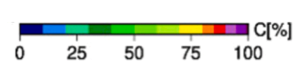


Fig. 6. SIC trends in Arctic waters (2004-2023).

shown in Fig. 5. Geospatially, severity accidents are primarily concentrated in the Norwegian Sea, Barents Sea, and along the coasts of Iceland and the United States. In contrast, pollution accidents are more dispersed and are predominantly located near coastal areas.

4. Evolution trends analysis

To further investigate the evolutionary trends of severity and pollution accidents, this section first provides a statistical analysis of accident frequencies over the past two decades and calculates the corresponding trends. The dataset is then evenly divided into four time periods (2004-2008, 2009-2013, 2014-2018, 2019-2023), and the evolution trends are analyzed from three perspectives: ship characteristics, accident types, and spatial distribution.

4.1. Arctic sea ice

Analyzing sea ice conditions in Arctic waters, using SIC as an

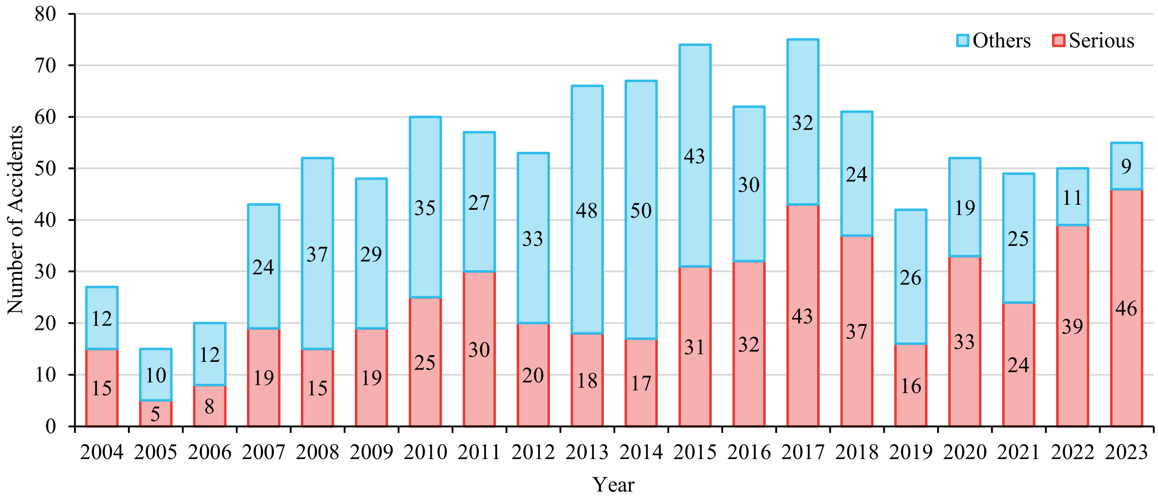
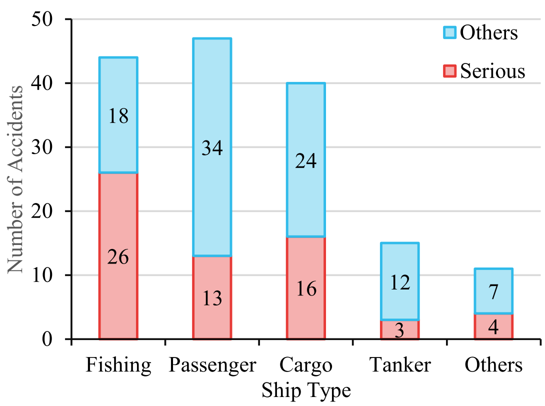
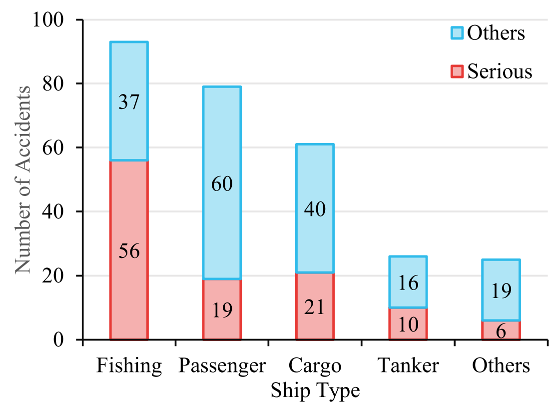


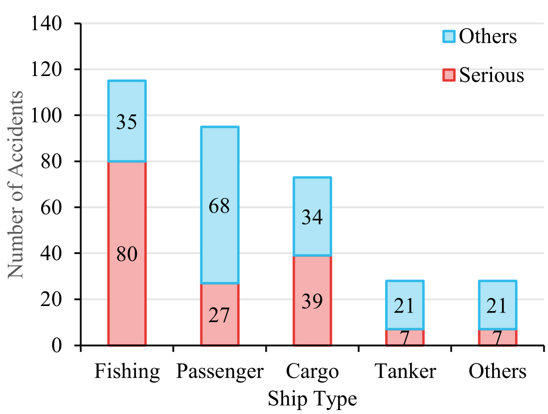
Fig. 7. Trends in the severity of consequences of accidents in Arctic waters.



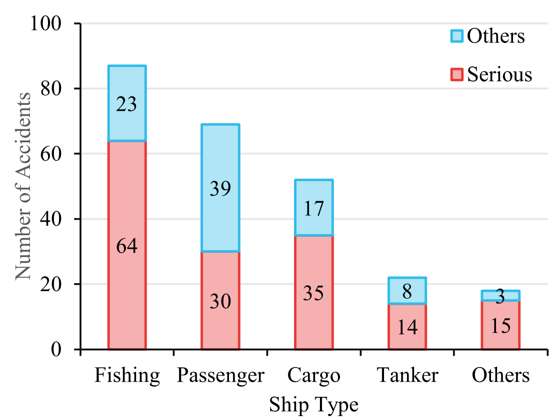
(a) 2004-2008.



(b) 2009-2013.



(c) 2014-2018.

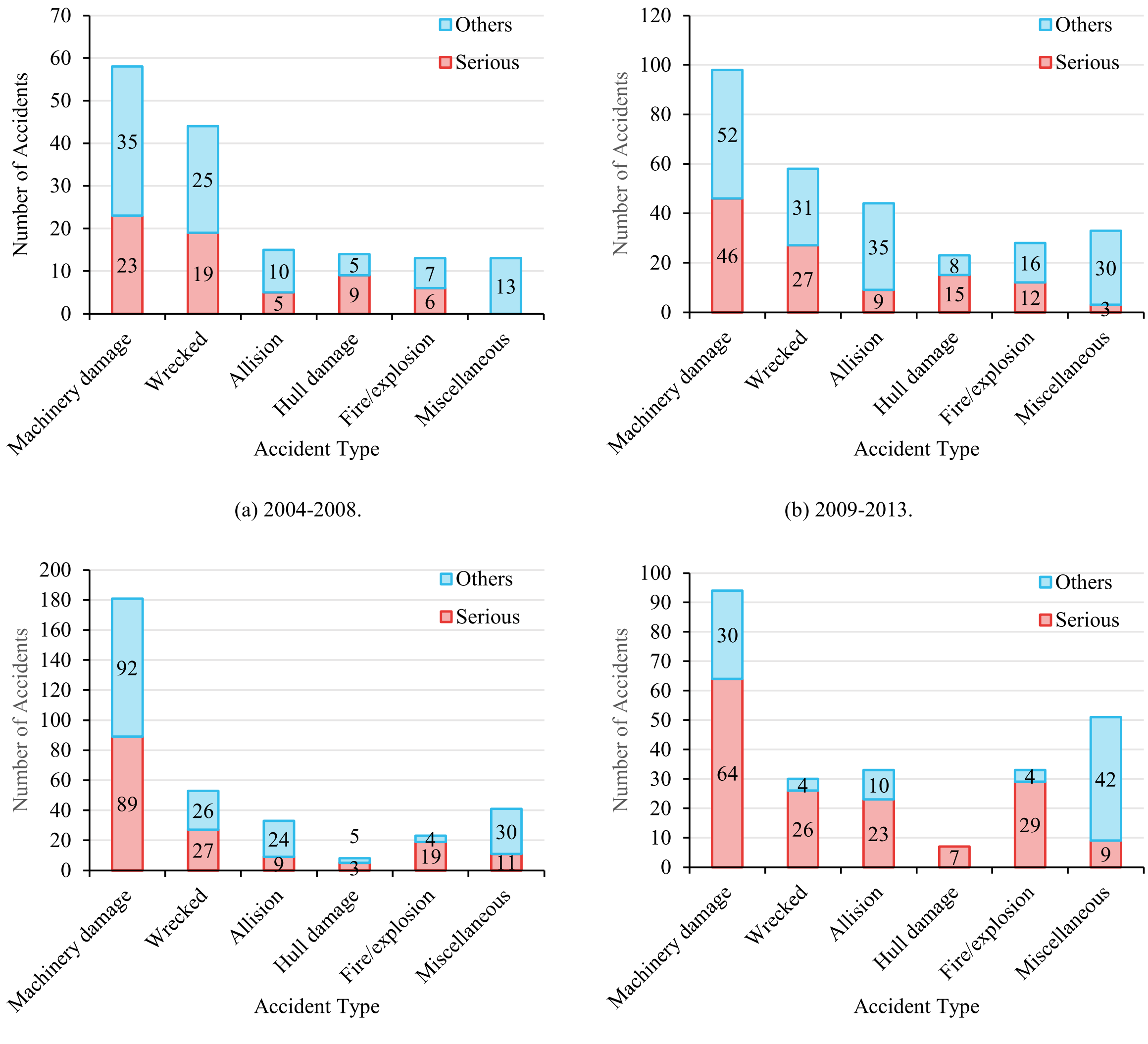


(d) 2019-2023.

Fig. 8. Trends in the severity of the consequences of accidents by ship type in Arctic waters.

example, involves importing SIC data for the summer (July 1) and winter (January 1) of 2004, 2014, and 2023 into the ArcGIS system, with the color map set to unique values. Fig. 6 illustrates the variation in the SIC over different periods.

Compared to winter, the extent of sea ice in the Arctic waters during summer significantly decreases. Over time, sea ice in the lower latitudes of the Arctic has gradually disappeared, while in the higher latitudes, both the extent and concentration of sea ice have also diminished. Although the extent of sea ice in the Arctic waters during winter has not changed significantly over time, the sea ice concentration in the lower latitudes has declined.



(a) 2004-2008.

(b) 2009-2013.

(c) 2014-2018.

(d) 2019-2023.

Fig. 9. Trends in the severity of consequences by accident type in Arctic waters.

4.2. *Severity accidents in Arctic waters*

As shown in Fig. 7, the total number of Arctic accidents shows a fluctuating upward trend between 2004 and 2023, peaking at 75 accidents in 2017. The number of other accidents shows an increasing and then decreasing trend, reaching a peak of 50 accidents in 2014. The number of severity accidents, on the other hand, show a steady upward trend, with the total number of severity accidents exceeding the total number of other accidents in 2016, peaking at 46 accidents in 2023. These results suggest that the probability of severity consequences from marine accidents in Arctic waters is increasing.

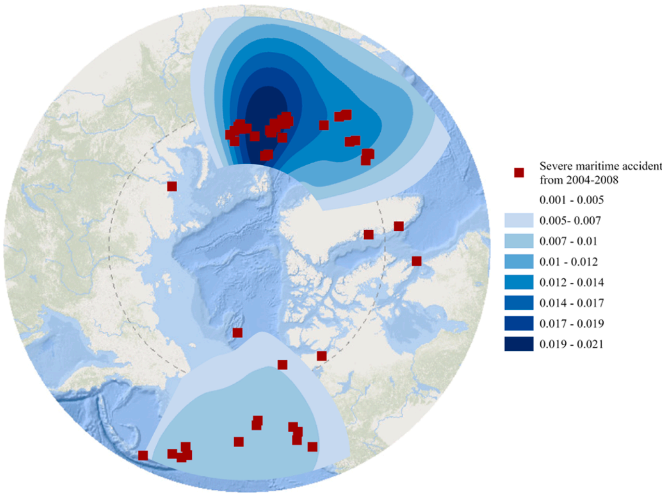
4.2.1. *Involved ships*

According to the results shown in Fig. 8, severity accidents across different ship types have exhibited varying degrees of upward trends. Fishing ships, represented by the largest sample size, consistently show a higher number of severity accidents than other ships, indicating that fishing ships in the Arctic have the highest probability of experiencing accidents with severe consequences. Although the number of other accidents involving passenger ships has always been higher than severity accidents, the proportion of severity accidents has significantly increased since the 2014-2018 period. For cargo ships, the number of severity accidents exceeded that of other accidents in the 2019-2023 period. Furthermore, tanker ships and other ship types, which have fewer samples, have also shown a trend in the past five years where the number of severity accidents exceeds that of other accidents.

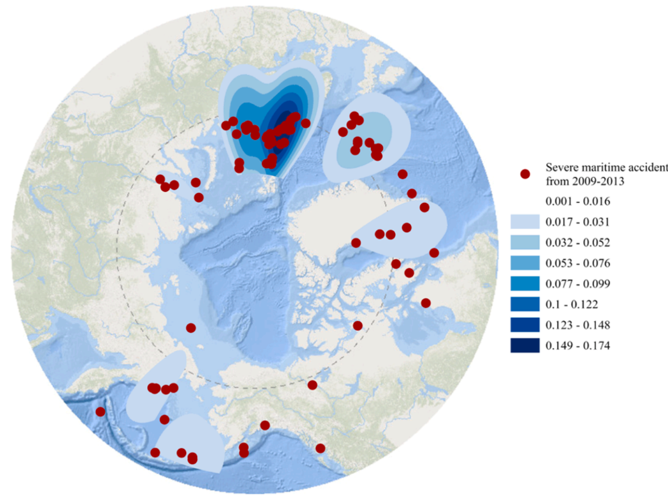
Overall, the total number of accidents involving ships was highest during the 2014-2018 period, while the proportion of severity accidents was highest during the 2019-2023 period. With the exception of passenger ships, the number of severity accidents for all other ship types has exceeded the number of other accidents. This indicates that the overall consequences of maritime accidents in the Arctic have become more severe.

4.2.2. *Accident type*

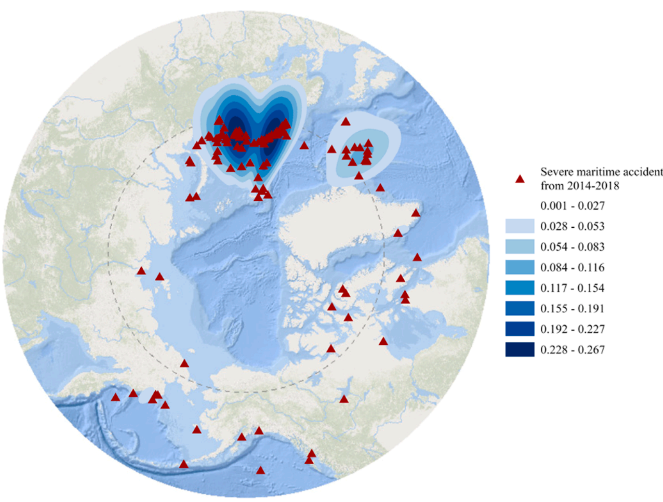
Fig. 9 shows the number of accidents for each time period for the six different accident types, with machinery damage being most common, followed by wrecked. The consequences of machinery damage, wrecked, allision, and fire/explosion accidents exhibit varying degrees of increased severity. Over the ten-year period from 2009 to 2018, the number of accidents significantly grew, with severity accidents in the 2019-2023 period far surpassing other accidents. Hull damage accidents, though occurring the least frequently, consistently have a severity accident proportion greater than 62%. Miscellaneous accidents show an increasing trend in frequency, with relatively few severity accidents, indicating a higher level of safety.



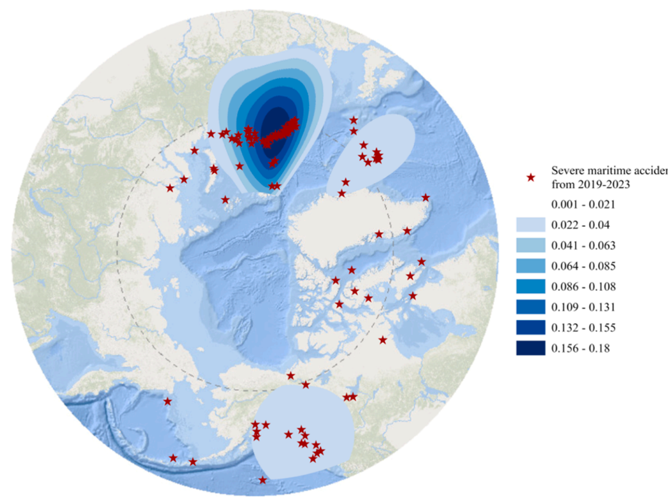
(a) 2004-2008.



(b) 2009-2013.



(c) 2014-2018.



(d) 2019-2023.

Fig. 10. Trends in the severity of accident consequences by region in Arctic waters.

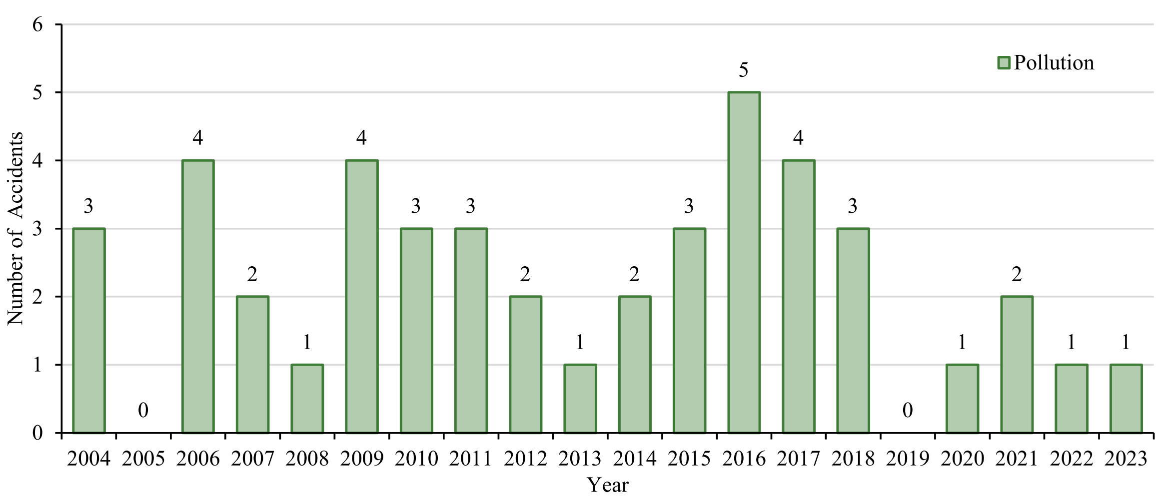


Fig. 11. Pollution trends of accidents in Arctic waters.

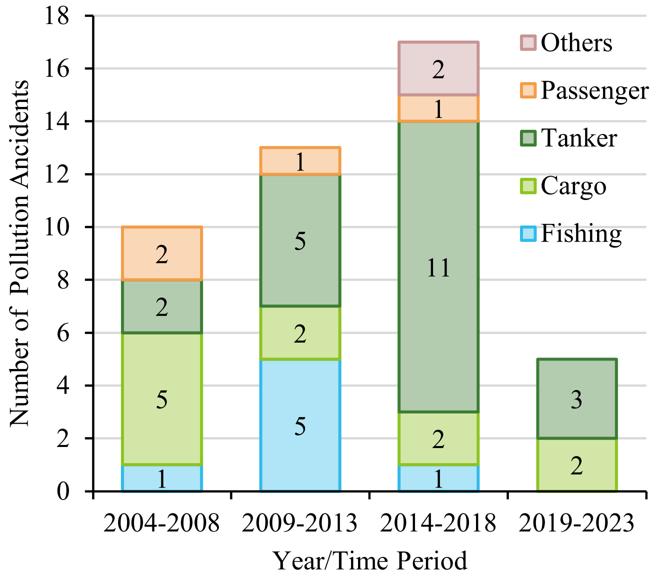


Fig. 12. Trends in the pollution consequences of accidents by ship type in Arctic waters.

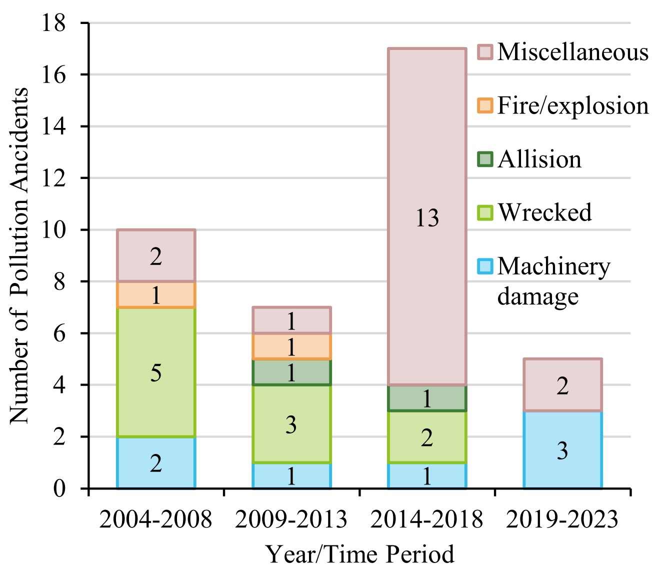


Fig. 13. Trends in the pollution consequences of accidents by type in Arctic waters.

4.2.3. Accident regions

The Norwegian Sea and Barents Sea have consistently been high-accident areas for accidents across all four periods. Over time, the frequency of accidents gradually increased in the waters around Iceland and along the northern coast of Canada.

As shown in Fig. 10, during the 2004-2008 period (a), maritime accidents were predominantly concentrated in the Norwegian Sea and Barents Sea, particularly along the Norwegian coast. In the subsequent 2009-2013 period (b), the spatial distribution of accidents expanded, extending into the Bering Sea and along the U.S. coast. By 2014-2018 (c), the frequency of severity accidents had markedly increased, with a pronounced concentration in the Norwegian Sea and Barents Sea. During the most recent 2019-2023 period (d), the spatial distribution of accidents became more dispersed, accompanied by a decline in accidents within the Bering Strait.

4.3. Pollution accidents in Arctic waters

According to Fig. 11, the probability of pollution accidents is low, with a general trend of growth followed by decline. On average, there are 2.25 accidents per year, with a maximum of five cases in 2016, followed by a steadily decreasing trend.

4.3.1. Involved ships

As shown in Fig. 12, among the five ship types, passenger and other types of ships are responsible for the fewest pollution accidents. Fishing ships experienced a relatively high number of pollution accidents between 2009 and 2013, with a total of five accidents. The frequency of pollution accidents involving cargo ships remained relatively stable. Due to the nature of their cargo, tanker ships caused the most pollution accidents, reaching a peak of 11 accidents during the 2014-2018 period.

4.3.2. Accident type

Fig. 13 presents the occurrence of pollution consequences for different accident types. Since hull damage accidents have no recorded pollution consequences, they are not included in the analysis. The accident type that caused the most pollution accidents was miscellaneous (24 cases), followed by wrecked (10 cases), with the majority of these accidents occurring during the 2014-2018 and 2004-2008 periods, respectively. Allision and fire/explosion accidents resulted in fewer pollution cases, with only two accidents recorded over the twenty-year period.

4.3.3. Accident regions

According to Fig. 14, Arctic ship pollution accidents occurred mainly near the Barents Sea, Iceland, and the North American coast during 2004-2008 and then spread to higher latitudes during 2009-2013, with a higher concentration in the Barents Sea. In the period 2014-2018, pollution accidents occurred mainly in the vicinity of the Northwest Passage. The number of pollution accidents then decreased dramatically, with the lowest number of accidents occurring during 2019-2023.

5. Influencing factors analysis

This section uses the integrated analysis-data dataset, which includes ship, accident, and environmental information. A bivariate probit model is constructed with accident severity and pollution as dependent variables to analyze the mechanisms through which various factors influence the outcomes of accidents.

5.1. Model description

A bivariate probit model was constructed with severity and pollution as dependent variables, and factors such as ship, accident, time, and environment as independent variables. The types of variables and their chi-square test results are listed in Table 7. A (*) indicates that a variable was set as a control variable.

Y1 and Y2 denote the dependent variables of severity and pollution, respectively, and in Table 7, C_Y1 and C_Y2 denote the results of the chi-square test of the independent variables Y1 and Y2. For ordered variables (year, GT, age, wind, etc.), the marginal variables are set as control variables, and for unordered variables (type of ship, type of accident, etc.), this study refers to the results of the chi-square test to set control variables.

5.2. Model estimation

The analysis-data dataset was used as a sample to construct a bivariate probit model for Arctic ship accidents by setting two dependent variables and 32 independent variables, as listed in Table 7. The model estimation results are shown in Table 8.

According to the model estimation results shown in Table 8, in terms

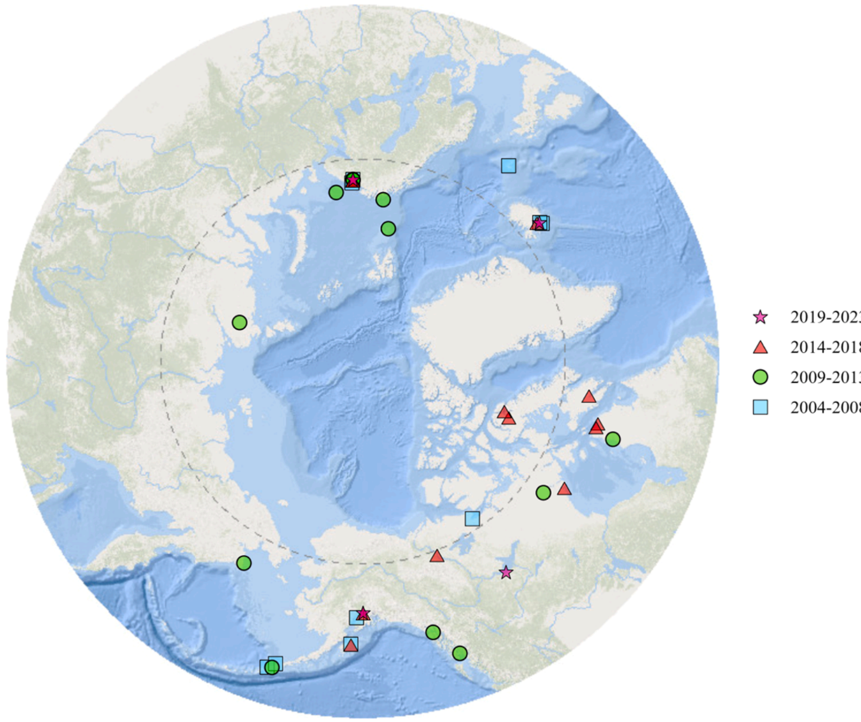


Fig. 14. The distribution of pollution accidents by region in Arctic waters.

of the overall fit of the model, the results of the Wald test showed that the hypothesis of rho = 0 was rejected (chi2(1) = 1648.66 with a p-value of 0.0000), indicating that there is a significant correlation between severity accidents (Y1) and pollution accidents (Y2) and that the bivariate probit model is appropriate. S_flag1 (Norway, Iceland, Denmark, or Finland) had a highly significant negative effect on pollution accidents (Coef. = -5.3224, $p < 0.001$), suggesting that it can reduce pollution accidents to some extent. In contrast, S_flag2 (Russia) and S_flag3 (Canada and U.S.A.) positively affect severity accidents and pollution accidents (Coef. = 0.3530 and 0.4637, p-value 0.0130 and 0.0630, respectively), indicating that these two types of flags increase to some extent the number of severity accidents and pollution accidents. This is attributed to Russia's frequent use of the Arctic routes, coupled with the prevalence of extreme weather conditions, rapidly changing ice conditions, and limited port and search-and-rescue resources in polar waters. In contrast, maritime activities in the Canada and USA regions are relatively scattered, with insufficient emergency response capabilities, increasing the likelihood of accidents escalating into severe consequences. Only S_type1 (fishing ship) and S_type4 (tanker ship) of the ship type variables significantly affect the consequences of accidents, with S_type1 positively affecting severity accidents (Coef. = 0.4624), and S_type4 having a more significant positive effect on pollution accidents (Coef. = 1.1653), indicating that fishing ships increase the occurrence of severity accidents and tanker ships increase the occurrence of accidents. The outcome is associated with vessel design and cargo characteristics. Fishing ships, typically small in tonnage and primarily designed for fishing operations, often lack advanced navigation and collision avoidance systems, making them less resistant to rough seas and more prone to severe accidents. In contrast, tanker ships, due to the hazardous nature of their cargo, may cause pollution even if the incident itself is not serious. GT did not have a significant effect on either accident outcome. Compared to ships older than 30 years, ships aged 10-20 years old increase the occurrence of severity accidents to some extent, while ships aged 20-30 years old decrease the occurrence of pollution accidents (Coef. = -0.5366).

When compared with the historical years 2004-2008, accidents during 2018-2023 significantly increase the probability of severity accidents (Coef. = 0.8599) and decrease the occurrence of pollution accidents (Coef. = -0.7960), which is in line with the conclusions reached in Section 4. In terms of accident types, all accident types had a significant impact on severity and pollution accidents. Compared to the miscellaneous category, all five accident types have a similar type of impact on consequences, with all having a positive impact on severity accidents (Coef. > 0) and a negative impact on pollution accidents (Coef. < 0). Machinery damage often result in loss of ship control, which in complex sea conditions can lead to secondary accidents (e.g., wrecked, allision et al). Allision, and hull damage are directly linked to casualties, though they do not necessarily involve pollutant cargo or fuel spills. Fire/explosion, being high-risk accidents, are directly associated with severe injuries or total vessel loss, thus exhibiting the highest positive impact coefficient (Coef. = 1.6326) on serious accidents.

For the navigational environmental factors, E_wind3 (above 13.9 m/s) significantly increases the occurrence of severity accidents compared to E_wind1 but does not have a significant effect on pollution accidents. Strong winds are typically accompanied by high-wave conditions, significantly impairing vessel maneuverability and navigation radar performance. This effect is particularly pronounced in Arctic waters, where the combination of low temperatures and strong winds can readily lead to severe maritime incidents. The SIC has a more significant negative effect on pollution accidents (Coef. = -0.7087), while the SIT has a significant positive effect on severity accidents (Coef. = 0.5127), indicating that the SIC decreases the occurrence of pollution accidents to some extent, while SIT increases the occurrence of severity accidents to some extent. High-SIC zones are generally identifiable, prompting vessels to adopt deliberate detour strategies to avoid areas with elevated ice concentrations as a safety precaution. Moreover, increased SIT imposes greater structural loads on vessel hulls, thereby substantially elevating the risk of serious marine accidents.

The estimation results indicate that the overall model is statistically significant, with a p-value of 0.0000, suggesting that the model as a whole is valid. The coefficient of the log-transformed correlation parameter /athrho is 0.2231 (p = 0.0360), corresponding to a ρ value of 0.2194. This implies a positive correlation between the error terms of the equations for accident severity and pollution.

Table 7
Variable descriptions.

Variables			Descriptions	Chi-square test	
				Mean	Std. Dev.~~2~~
Independent variables					
Ship	Flag	S_flag1	1: the flag is Norway, Iceland, Denmark, or Finland, 0: otherwise.	0.31	0.46
		S_flag2	1: the flag is Russia, 0: otherwise.	0.25	0.43
		S_flag3	1: the flag is Canada or U.S.A., 0: otherwise.	0.18	0.38
		S_flag*	1: the flag is Others (non-Arctic countries), 0: otherwise.	-	-
	GT	*S_GT1*	1: the GT of the ship is (0,3000] tons, 0: otherwise.	0.28	0.45
		S_GT2	1: the GT of the ship is (3000,10000] tons, 0: otherwise.	0.26	0.44
		S_GT3	1: the GT of the ship is (10000,20000] tons, 0: otherwise.	0.25	0.43
		*S_GT4**	1: the GT of the ship is (20000, $+\infty$) tons, 0: otherwise.	-	-
	Age	*S_age1*	1: the age of the ship is (0,10] years, 0: otherwise.	0.22	0.42
		S_age2	1: the age of the ship is (10,20] years, 0: otherwise.	0.20	0.40
		S_age3	1: the age of the ship is (20,30] years, 0: otherwise.	0.26	0.44
		*S_age**	1: the age of the ship is (30, $+\infty$) years, 0: otherwise.	-	-
	Type	S_type1	1: the type of the ship is fishing ship, 0: otherwise.	0.33	0.47
		S_type2	1: the type of the ship is passenger ship, 0: otherwise.	0.28	0.45
		S_type3	1: the type of the ship is cargo ship, 0: otherwise.	0.22	0.41
		S_type4	1: the type of the ship is tanker ship, 0: otherwise.	0.09	0.28
		S_type*	1: the type of the ship is others, 0: otherwise.	-	-
Accident	Type	*A_type1*	1: accident type is machinery damage, 0: otherwise.	0.42	0.49
		A_type2	1: accident type is wrecked, 0: otherwise.	0.18	0.38
		A_type3	1: accident type is allision (collision, contact), 0: otherwise.	0.12	0.33
		A_type4	1: accident type is hull damage, 0: otherwise.	0.09	0.29
		A_type5	1: accident type is fire/explosion, 0: otherwise.	0.05	0.22
		*A_type**	1: accident type is miscellaneous, 0: otherwise.	-	-
	Year	*A_year2*	1: 2009-2013, 0: otherwise.	0.28	0.45
		A_year3	1: 2014-2018, 0: otherwise.	0.33	0.47
		A_year4	1: 2019-2023, 0: otherwise.	0.24	0.43
		*A_year**	1: 2004-2008, 0: otherwise.	-	-
Environment	SIC	*E_SIC*	1: [0.1, 1) % 0: (0, 0.1) %	0.08	0.26
	SIT	*E_SIT*	1: above 10 cm 0: below 10 cm	0.03	0.18
	Wind state	*E_Wind1**	1: [0,8) m/s, 0: otherwise.	-	-
		E_Wind2	1: [8,13.9) m/s, 0: otherwise.	0.37	0.48
		E_Wind3	1: [13.9, $+\infty$) m/s, 0: otherwise.	0.16	0.37
Dependent variables					
Consequence	Severity	*Y1*	1: severity (including total loss) 0: others	0.48	0.50
	Pollution	*Y2*	1: pollution 0: others	0.04	0.20

Furthermore, the Wald test rejects the null hypothesis of no correlation at the 5% significance level, providing empirical evidence of a significant association between accident severity and pollution occurrence. These results support the use of the bivariate probit model, which more effectively captures the underlying interdependence between the two outcome variables compared to estimating two separate univariate probit models.

6. Discussion

The relationship between the control variable and the two observed dependent variables is nonlinear, so the coefficients themselves do not directly reflect the actual effect of the independent variable on the probability of the dependent variable. In order to investigate the amount of change in the probability of the dependent variable, taking a value of 1 for each unit increase in the independent variable, this section calculates the marginal effects shown in Table 9.

The results show a significant effect of the accident type. For severity accidents (Y1), the marginal effects of A_type1, A_type2, A_type3, A_type4, and A_type5 are all significantly positive (marginal values of 0.3714, 0.4245, 0.2696. 0.4453, and 0.5289, respectively) and have a significance level of 1%, indicating that the marginal effects of these variables on Y1 are significantly positive. For example, the marginal effect of A_type1 is 0.3714, implying that, all else being equal, the probability of Y1 will increase by 37.14% for each unit increase in A_type1. For pollution accidents (Y2), the marginal effects for all five accident types are significantly negative (the marginal results are -0.0673, -0.0337, -0.0821, -0.0846, and -0.0984, respectively) and have a significance level of 1%, suggesting that a one-unit increase in all five accident types will decrease the probability of Y2 by 3% to 10%.

For the ship type, S_type1 has a positive marginal effect on Y1 (marginal value $= 0.1498$, $P = 0.0150$), indicating that each unit increase in fishing ships will result in a 14.98% increase in the probability of a severity accidents. S_type4 has a significant effect on pollution accidents, with a marginal effect of 0.0727, indicating that each unit increase in tanker ships will result in a 7.27% increase in the probability of pollution accidents.

As for the environmental factors, strong wind (a wind speed over 13.9m/s) has a significant positive effect on severity accidents, and each unit increase in the strong wind will increase the probability of severity accidents by 7.25%. The SIC has a negative effect on pollution accidents, and each unit increase in the SIC decreases the probability of pollution accidents by 4.42%. Meanwhile, the SIT has a positive effect on severity accidents, with each unit increase resulting in a 16.61% increase in the probability of severity accidents.

7. Conclusion

This paper has proposed a data-driven framework to identify the evolution trends and influencing factors of maritime accidents in Arctic

Table 8
Estimated parameters of the bivariate probit model for the maritime accidents in Arctic waters.

Variables	Explanation	Y1				Y2			
		Coef.	Robust Std. Err.	z	P>|z|	Coef.	Robust Std. Err.	z	P>|z|
S_flag1	Norway, Iceland, Denmark, or Finland	-0.0516	0.1242	-0.42	0.6780	-5.3224	0.29	-18.0700	0.0000***
S_flag2	Russia	0.3530	0.1419	2.49	0.0130**	0.3705	0.24	1.5200	0.1280
S_flag3	Canada	0.0202	0.1570	0.13	0.8980	0.4637	0.25	1.8600	0.0630*
S_type1	Fishing	0.4624	0.1919	2.41	0.0160**	0.2345	0.39	0.6100	0.5430
S_type2	Passenger	-0.1695	0.1904	-0.89	0.3730	-0.1658	0.46	-0.3600	0.7170
S_type3	Cargo ship	0.2615	0.2033	1.29	0.1980	0.4677	0.38	1.2400	0.2160
S_type4	Tanker ship	-0.0085	0.2222	-0.04	0.9690	1.1653	0.38	3.1100	0.0020***
S_GT1	(0,3000] tons	-0.0247	0.1553	-0.16	0.8730	-0.2452	0.36	-0.6700	0.5000
S_GT2	(3000,10000] tons	0.0002	0.1468	0.00	0.9990	0.0561	0.29	0.2000	0.8450
S_GT3	(10000,20000] tons	-0.1589	0.1357	-1.17	0.2420	0.1239	0.23	0.5300	0.5960
S_age1	(0,10] years	-0.1790	0.1326	-1.35	0.1770	-0.0240	0.29	-0.0800	0.9340
S_age2	(10,20] years	0.2517	0.1299	1.94	0.0530*	0.2179	0.27	0.8200	0.4140
S_age3	(20,30] years	-0.0137	0.1139	-0.12	0.9040	-0.5366	0.25	-2.1300	0.0330**
A_year2	2009-2013	0.0214	0.1369	0.16	0.8760	-0.1966	0.26	-0.7600	0.4490
A_year3	2014-2018	0.2050	0.1354	1.51	0.1300	-0.3193	0.24	-1.3400	0.1820
A_year4	2019-2023	0.8599	0.1458	5.90	0.0000***	-0.7960	0.30	-2.6300	0.0090***
A_type1	Machinery damage	1.1465	0.1631	7.03	0.0000***	-1.0794	0.22	-4.9000	0.0000***
A_type2	Wrecked	1.3103	0.1806	7.26	0.0000***	-0.5395	0.26	-2.0900	0.0370**
A_type3	Allision	0.8322	0.1928	4.32	0.0000***	-1.3162	0.37	-3.5200	0.0000***
A_type4	Hull damage	1.3744	0.2069	6.64	0.0000***	-1.3560	0.42	-3.2300	0.0010***
A_type5	Fire/explosion	1.6326	0.2416	6.76	0.0000***	-1.5781	0.45	-3.5100	0.0000***
E_wind2	[8, 13.9) m/s	0.0756	0.0956	0.79	0.4290	-0.1098	0.18	-0.6000	0.5490
E_wind3	[13.9, +∞) m/s	0.2239	0.1267	1.77	0.0770*	-0.2288	0.28	-0.8100	0.4180
E_SIC	[0.1, 1) %	0.0080	0.1650	0.05	0.9610	-0.7087	0.40	-1.7600	0.0790*
E_SIT	Above 10 cm	0.5127	0.2438	2.10	0.0350**	0.3190	0.38	0.8400	0.4010
_cons		-1.6541	0.3000	-5.51	0.0000***	-0.9961	0.53	-1.8900	0.0580*
/athrho		0.2231	0.1066	2.09	0.0360	0.2231	0.1066	2.09	0.0360
P		0.2194	0.1014			0.2194	0.1014		

Number of obs. = 1028
Log pseudolikelihood = -700.7198
Wald test of rho=0: chi2(1) = 4.3821
Prob > chi2 = 0. 0.0363
Note: ***, **, * ==> Significance at 1%, 5%, 10% level.

Table 9
Marginal effect calculation results.

Variables	Explanation	Y1				Y2			
		dy/dx	Std. Err.	z	P>|z|	dy/dx	Std. Err.	z	P>|z|
S_flag1	Norway etc.	-0.0167	0.0402	-0.42	0.6770	-0.3320	0.0440	-7.54	0.0000***
S_flag2	Russia	0.1144	0.0456	2.51	0.0120**	0.0231	0.0151	1.53	0.1270
S_flag3	Canada	0.0065	0.0509	0.13	0.8980	0.0289	0.0160	1.80	0.0710*
S_type1	Fishing	0.1498	0.0617	2.43	0.0150**	0.0146	0.0241	0.61	0.5440
S_type2	Passenger	-0.0549	0.0616	-0.89	0.3730	-0.0103	0.0286	-0.36	0.7170
S_type3	Cargo ship	0.0847	0.0658	1.29	0.1980	0.0292	0.0237	1.23	0.2190
S_type4	Tanker ship	-0.0028	0.0720	-0.04	0.9690	0.0727	0.0233	3.12	0.0020***
S_GT1	(0,3000] tons	-0.0080	0.0503	-0.16	0.8730	-0.0153	0.0227	-0.67	0.5000
S_GT2	(3000,10000] tons	0.0001	0.0476	0.00	0.9990	0.0035	0.0179	0.20	0.8450
S_GT3	(10000,20000] tons	-0.0515	0.0438	-1.17	0.2400	0.0077	0.0145	0.53	0.5940
S_age1	(0,10] years	-0.0580	0.0429	-1.35	0.1760	-0.0015	0.0180	-0.08	0.9340
S_age2	(10,20] years	0.0816	0.0418	1.95	0.0510*	0.0136	0.0167	0.81	0.4160
S_age3	(20,30] years	-0.0044	0.0369	-0.12	0.9040	-0.0335	0.0162	-2.07	0.0390**
A_year2	2009-2013	0.0069	0.0443	0.16	0.8760	-0.0123	0.0162	-0.76	0.4480
A_year3	2014-2018	0.0664	0.0437	1.52	0.1290	-0.0199	0.0153	-1.30	0.1930
A_year4	2019-2023	0.2786	0.0450	6.20	0.0000***	-0.0497	0.0194	-2.56	0.0100***
A_type1	Machinery damage	0.3714	0.0485	7.66	0.0000***	-0.0673	0.0136	-4.97	0.0000***
A_type2	Wrecked	0.4245	0.0537	7.91	0.0000***	-0.0337	0.0157	-2.15	0.0320***
A_type3	Allision	0.2696	0.0607	4.44	0.0000***	-0.0821	0.0233	-3.53	0.0000***
A_type4	Hull damage	0.4453	0.0623	7.15	0.0000***	-0.0846	0.0255	-3.31	0.0010***
A_type5	Fire/explosion	0.5289	0.0730	7.25	0.0000***	-0.0984	0.0298	-3.31	0.0010***
E_wind2	[8, 13.9) m/s	0.0245	0.0309	0.79	0.4290	-0.0068	0.0114	-0.60	0.5480
E_wind3	[13.9, +∞) m/s	0.0725	0.0409	1.77	0.0760*	-0.0143	0.0174	-0.82	0.4120
E_SIC	[0.1, 1) %	0.0026	0.0535	0.05	0.9610	-0.0442	0.0248	-1.78	0.0750*
E_SIT	Above 10 cm	0.1661	0.0788	2.11	0.0350**	0.0199	0.0241	0.83	0.4090

waters, considering both accident severity and pollution. By integrating multi-source maritime accident and sea ice environmental data, a bivariate probit model has been developed to explore the factors affecting the coupled relationship between accident severity and pollution. Additionally, a marginal effect analysis has been conducted to quantify the impact of these factors on accident severity and pollution. The results indicate a significant upward trend in the incidence of severity maritime accidents in Arctic waters, with the accident distribution exhibiting clear spatial clustering. These accidents are mainly concentrated in the Norwegian Sea, the Barents Sea, and the waters around Iceland and the northern coast of Canada, with a clear tendency to spread to higher latitudes. In terms of the distribution of ship types, oil tankers are the main ship type causing pollution accidents, while fishing ships are the main ship type causing severity accidents. Analyses based on bivariate probit models and their marginal effects showed that the flag state, ship type, accident type, and environmental factors had a significant effect on the occurrence of both severity accidents and pollution accidents. In terms of ship factors, a Russian flag was strongly associated with the occurrence of severity accidents, while Nordic and Canadian flags were strongly associated with the occurrence of pollution accidents. Fishing ships significantly increased the probability of severity accidents, while tanker ships increased the probability of pollution accidents. Newer ships were able to reduce the probability of severity accidents and pollution accidents to a certain extent, relative to older ships. In terms of accident factors, mechanical damage, wrecked, allision, hull damage, and fire/explosion significantly increased the probability of severity accidents, but these factors have different degrees of negative impact on the probability of pollution accidents. Among environmental factors, strong winds (wind speed > 13.9m/s) and increased SIC significantly increased the probability of severity accidents, while sea ice with a thickness greater than 10 cm significantly increased the probability of severity accidents. This suggests that ice concentration reflects the complexity and dynamics of the ice environment, and that a high ice concentration increases the uncertainty of ship navigation and increases the probability of accidents. Meanwhile, SIT reflects the physical strength of the ice, and thicker ice areas are usually regarded as high-risk areas, where ships will actively avoid or adopt safer navigation strategies. These findings provide an important empirical basis for understanding the main factors influencing ship accidents in Arctic waters and provide scientific support for the development of risk management and environmental protection policies for Arctic shipping.

Future research could further refine the classification and quantitative analysis of environmental variables such as sea ice using real data to explore the differentiated impact of various ice conditions and other meteorological factors on accident outcomes. Additionally, considering environmental changes in the Arctic under the context of climate change and expanding the research dimensions will provide more scientific references for the formulation of maritime safety strategies in Arctic waters.

CRediT authorship contribution statement

Shanshan Fu: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Mengfei Cui:** Writing – review & editing, Software, Methodology, Data curation. **Ningji Wu:** Writing – review & editing, Writing – original draft, Software, Data curation. **Mingyang Zhang:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Xiao Lang:** Writing – review & editing. **Wengang Mao:** Writing – review & editing, Validation, Supervision.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, 'Evolution Trends and Influencing Factors Analysis for the Severity and Pollution of Maritime Accidents in Arctic Waters from Multi-Source Data'.

Acknowledgment

This study is supported by the National Natural Science Foundation of China under grants 52271363, and Shanghai Rising-Star Program 22QC1400600.

Data availability

Data will be made available on request.

References

[1] Theocharis D, Pettit S, Rodrigues VS, Haider J. Arctic shipping: a systematic literature review of comparative studies. J Transp Geogr 2018;69:112–28.
[2] Ng A, Andrews J, Babb D, Lin YF, Becker A. Implications of climate change for shipping: opening the Arctic seas. Wiley Interdiscip Rev.-Clim Change 2018;9(2).
[3] Kondratenko AA, Kujala P, Hirdaris SE. Holistic and sustainable design optimization of Arctic ships. Ocean Eng 2023;275:114095.
[4] Hermann RR, Lin N, Lebel J, Kovalenko A. Arctic transshipment hub planning along the Northern Sea Route: a systematic literature review and policy implications of Arctic port infrastructure. Mar Policy 2022;145:105275.
[5] Liu Y, Ma X, Qiao W, Ma L, Han B. A novel methodology to model disruption propagation for resilient maritime transportation systems–a case study of the Arctic maritime transportation system. Reliab Eng Syst Saf 2024;241:109620.
[6] Kandel R, Baroud H. A data-driven risk assessment of Arctic maritime incidents: using machine learning to predict incident types and identify risk factors. Reliab Eng Syst Saf 2024;243:109779.
[7] Browne T, Taylor R, Veitch B, Helle I, Parviainen T, Khan F, Smith D. A general method to combine environmental and life-safety consequences of Arctic ship accidents. Saf Sci 2022;154:105855.
[8] Fu S, Tang Q, Zhang M, Han B, Wu Z, Mao W. A data-driven framework for risk and resilience analysis in maritime transportation systems: A case study of domino effect accidents in Arctic waters. Reliab Eng Syst Saf 2025;260:111049.
[9] Kum S, Sahin B. A root cause analysis for Arctic marine accidents from 1993 to 2011. Saf Sci 2015;74:206–20.
[10] Montewka J, Goerlandt F, Kujala P, Lensu M. Towards probabilistic models for the prediction of a ship performance in dynamic ice. Cold Reg Sci Technol 2015;112:14–28.
[11] Fu S, Zhang D, Montewka J, Zio E, Yan X. A quantitative approach for risk assessment of a ship stuck in ice in Arctic waters. Saf Sci 2018;107:145–54.
[12] Vanhatalo J, Huuhtanen J, Bergström M, Helle I, Mäkinen J, Kujala P. Probability of a ship becoming beset in ice along the Northern Sea Route - a Bayesian analysis of real-life data. Cold Reg Sci Technol 2021;184:103238.
[13] Khan B, Khan F, Veitch B. A dynamic Bayesian network model for ship-ice collision risk in the Arctic waters. Saf Sci 2020;130:104858.
[14] Yu Y, Liu K, Fu S, Chen J. Framework for process risk analysis of maritime accidents based on resilience theory: a case study of grounding accidents in Arctic waters. Reliab Eng Syst Saf 2024;249:110202.
[15] Fu S, Yu Y, Chen J, Xi Y, Zhang M. A framework for quantitative analysis of the causation of grounding accidents in Arctic shipping. Reliab Eng Syst Saf 2022;226:108706.
[16] Zhang M, Zhang D, Goerlandt F, Yan X, Kujala P. Use of HFACS and fault tree model for collision risk factors analysis of icebreaker assistance in ice-covered waters. Saf Sci 2019;111:128–43.
[17] Fu S, Zhang D, Montewka J, Yan X, Zio E. Towards a probabilistic model for predicting ship besetting in ice in Arctic waters. Reliab Eng Syst Saf 2016;155:124–36.
[18] Xu S, Kim E, Haugen S, Zhang M. A Bayesian network risk model for predicting ship besetting in ice during convoy operations along the Northern Sea Route. Reliab Eng Syst Saf 2022;223:108475.
[19] Khan B, Khan F, Veitch B, Yang M. An operational risk analysis tool to analyze marine transportation in Arctic waters. Reliab Eng Syst Saf 2018;169:485–502.
[20] Liu C, Kulkarni K, Suominen M, Kujala P, Musharraf M. On the data-driven investigation of factors affecting the need for icebreaker assistance in ice-covered waters. Cold Reg Sci Technol 2024;221:104173.
[21] Zhu X, Hu S, Li Z, Wu J, Yang X, Fu S, Han B. Risk performance analysis approach for convoy operations via a hybrid model of STPA and DBN: a case from ice-covered waters. Ocean Eng 2024;302:117570.
[22] Fu S, Zhang Y, Zhang M, Han B, Wu Z. An object-oriented Bayesian network model for the quantitative risk assessment of navigational accidents in ice-covered Arctic waters. Reliab Eng Syst Saf 2023;238:109459.
[23] Fu S, Goerlandt F, Xi Y. Arctic shipping risk management: a bibliometric analysis and a systematic review of risk influencing factors of navigational accidents. Saf Sci 2021;139:105254.
[24] Zhang W, Xiao Y, Liu C, Zhang M, Wang L, Feng L. A kinematic model for collaborative icebreaker convoy operations in ice-covered waters. Ocean Eng 2024;304:117870.
[25] Wang H, Liu Z, Liu Z, Wang X, Wang J. GIS-based analysis on the spatial patterns of global maritime accidents. Ocean Eng 2022;245:110569.
[26] Wang J, Zhou Y, Zhuang L, Shi L, Zhang S. Study on the critical factors and hot spots of crude oil tanker accidents. Ocean Coast Manag 2022;217:106010.

[27] Zhang Y, Sun X, Chen J, Cheng C. Spatial patterns and characteristics of global maritime accidents. Reliab Eng Syst Saf 2021;206:107310.
[28] Feng Y, Wang X, Chen Q, Yang Z, Wang J, Li H, et al. Prediction of the severity of marine accidents using improved machine learning. Transp Res E: Logist Transp Rev 2024;188:103647.
[29] Wang H, Liu Z, Wang X, Graham T, Wang J. An analysis of factors affecting the severity of marine accidents. Reliab Eng Syst Saf 2021;210:107513.
[30] Li H, Zhou K, Zhang C, Bashir M, Yang Z. Dynamic evolution of maritime accidents: comparative analysis through data-driven Bayesian networks. Ocean Eng 2024; 303:117736.
[31] Zhang Y, Zhai Y, Fu S, Shi M, Jiang X. Quantitative analysis of maritime piracy at global and regional scales to improve maritime security. Ocean Coast Manag 2024; 248:106968.
[32] Sui Z, Wen Y, Huang Y, Song R, Piera MA. Maritime accidents in the Yangtze River: a time series analysis for 2011–2020. Accid Anal Prev 2023;180:106901.
[33] Çakır E, Fışkın R, Sevgili C. Investigation of tugboat accidents severity: an application of association rule mining algorithms. Reliab Eng Syst Saf 2021;209: 107470.
[34] Zhou X, Ruan X, Wang H, Zhou G. Exploring spatial patterns and environmental risk factors for global maritime accidents: a 20-year analysis. Ocean Eng 2023;286: 115628.
[35] Li H, Yang Z. Towards safe navigation environment: the imminent role of spatio-temporal pattern mining in maritime piracy incidents analysis. Reliab Eng Syst Saf 2023;238:109422.
[36] Lau Y, Yang Z, Yin J, Lei Z, Ching-Pong Poo M. Assessing vessel pollution risk in Asian areas: a comparative analysis based on data-driven Bayesian network approach. Ocean Coast Manag 2025;262:107549.
[37] Cao Y, Iulia M, Majumdar A, Feng Y, Xin X, Wang X, et al. Investigation of the risk influential factors of maritime accidents: a novel topology and robustness analytical framework. Reliab Eng Syst Saf 2025;254:110636.
[38] Chen J, Di Z, Shi J, Shu Y, Wan Z, Song L, et al. Marine oil spill pollution causes and governance: a case study of Sanchi tanker collision and explosion. J Clean Prod 2020;273:122978.
[39] Wan S, Yang X, Chen X, Qu Z, An C, Zhang B, et al. Emerging marine pollution from container ship accidents: risk characteristics, response strategies, and regulation advancements. J Clean Prod 2022;376:134266.
[40] Özaydın E, Fışkın R, Uğurlu Ö, Wang J. A hybrid model for marine accident analysis based on Bayesian network (BN) and association rule mining (ARM). Ocean Eng 2022;247:110705.
[41] Gan L, Gao Z, Zhang X, Xu Y, Liu RW, Xie C, et al. Graph neural networks enabled accident causation prediction for maritime vessel traffic. Reliab Eng Syst Saf 2025; 257:110804.
[42] Liu C, Suominen M, Musharraf M. An ensemble machine learning model for predicting the need for icebreaker assistance in ice-covered waters. Eng Appl Artif Intell 2025;158:111489.
[43] Lan H, Ma X, Ma L, Qiao W. Pattern investigation of total loss maritime accidents based on association rule mining. Reliab Eng Syst Saf 2023;229:108893.
[44] Chen J, Bian W, Wan Z, Yang Z, Zheng H, Wang P. Identifying factors influencing total-loss marine accidents in the world: analysis and evaluation based on ship types and sea regions. Ocean Eng 2019;191:106495.
[45] Huang C, Hu S. Factors correlation mining on maritime accidents database using association rule learning algorithm. Cluster Comput 2019;22(2):4551–9.
[46] Shi J, Liu Z, Feng Y, Wang X, Zhu H, Yang Z, et al. Evolutionary model and risk analysis of ship collision accidents based on complex networks and dematel. Ocean Eng 2024;305:117965.
[47] Sevgili C, Fiskin R, Cakir E. A data-driven Bayesian network model for oil spill occurrence prediction using tankship accidents. J Clean Prod 2022;370:133478.
[48] Chen J, Zhang W, Li S, Zhang F, Zhu Y, Huang X. Identifying critical factors of oil spill in the tanker shipping industry worldwide. J Clean Prod 2018;180:1–10.
[49] Wang K, Yuan Y, Chen M, Wang D. A POIs based method for determining spatial distribution of urban fire risk. Process Saf Environ Prot 2021;154:447–57.
[50] Christodoulou A, Dalaklis D, Raneri P, Sheehan R. An overview of the legal search and rescue framework and related infrastructure along the Arctic Northeast Passage. Mar Policy 2022;138:104985.
[51] Stipancic J, Zangenehpour S, Miranda-Moreno L, Saunier N, Granié M. Investigating the gender differences on bicycle-vehicle conflicts at urban intersections using an ordered logit methodology. Accid Anal Prev 2016;97:19–27.
[52] Chiou Y, Hwang C, Chang C, Fu C. Modeling two-vehicle crash severity by a bivariate generalized ordered probit approach. Accid Anal Prev 2013;51:175–84.
[53] Spreen G, Kaleschke L, Heygster G. Sea ice remote sensing using AMSR-E 89-GHz channels. J Geophy Res Atmos 2008;113(C2):C02S03.
[54] Paţilea C, Heygster G, Huntemann M, Spreen G. Combined SMAP–SMOS thin sea ice thickness retrieval. Cryosphere 2019;13(2):675–91.
[55] Hersbach H, Bell B, Berrisford P, Hirahara S, Horanyi A, Munoz-Sabater J, et al. The ERA5 global reanalysis. Q J R Meteorol Soc 2020;146(730):1999–2049.
[56] Huang D, Liang T, Hu S, Loughney S, Wang J. Characteristics analysis of intercontinental sea accidents using weighted association rule mining: evidence from the Mediterranean Sea and Black Sea. Ocean Eng 2023;287:115839.
[57] Marino M, Cavallaro L, Castro E, Musumeci RE, Martignoni M, Roman F, et al. New frontiers in the risk assessment of ship collision. Ocean Eng 2023;274:113999.
[58] Chang W, Ung S, Hu H. A marine accident analysis based on data-driven Bayesian network considering weather conditions and its application to Taiwanese waters. Ocean Eng 2024;309:118527.
[59] Jiang H, Zhang J, Wan C, Zhang M, Soares CG. A data-driven Bayesian network model for risk influencing factors quantification based on global maritime accident database. Ocean Coast Manag 2024;259:107473.