




Article

Analysis of Factors Affecting the Effectiveness of Oil Spill Clean-Up: A Bayesian Network Approach

Liangxia Zhong^{1,2,3}, Jiaxin Wu^{1,2,3} , Yiqing Wen^{1,2,3} , Bingjie Yang⁴, Manel Grifoll⁵ , Yunping Hu^{1,2,3,*} and Pengjun Zheng^{1,2,3,*}

¹ Faculty of Maritime and Transportation, Ningbo University, Ningbo 315832, China

² Jiangsu Province Collaborative Innovation Center for Modern Urban Traffic Technologies, Nanjing 211189, China

³ National Traffic Management Engineering & Technology Research Centre, Ningbo University Sub-Center, Ningbo 315832, China

⁴ Ningbo Development Planning Institute, Ningbo 315832, China

⁵ Barcelona School of Nautical Studies, Universitat Politècnica de Catalunya (UPC—BarcelonaTech), 08003 Barcelona, Spain

* Correspondence: huyunping@nbu.edu.cn (Y.H.); zhengpengjun@nbu.edu.cn (P.Z.)

Abstract: Ship-related marine oil spills pose a significant threat to the environment, and while it may not be possible to prevent such incidents entirely, effective clean-up efforts can minimize their impact on the environment. The success of these clean-up efforts is influenced by various factors, including accident-related factors such as the type of accident, location, and environmental weather conditions, as well as emergency response-related factors such as available resources and response actions. To improve targeted and effective responses to oil spills resulting from ship accidents and enhance oil spill emergency response methods, it is essential to understand the factors that affect their effectiveness. In this study, a data-driven Bayesian network (TAN) analysis approach was used with data from the U.S. Coast Guard (USCG) to identify the key accident-related factors that impact oil spill clean-up performance. The analysis found that the amount of discharge, severity, and the location of the accident are the most critical factors affecting the clean-up ratio. These findings are significant for emergency management and planning oil spill clean-up efforts.

Keywords: oil spill clean-up ratio; analysis of factors; Bayesian network



Citation: Zhong, L.; Wu, J.; Wen, Y.; Yang, B.; Grifoll, M.; Hu, Y.; Zheng, P. Analysis of Factors Affecting the Effectiveness of Oil Spill Clean-Up: A Bayesian Network Approach.

Sustainability **2023**, *15*, 4965.

<https://doi.org/10.3390/su15064965>

Academic Editor: Tim Gray

Received: 1 February 2023

Revised: 4 March 2023

Accepted: 8 March 2023

Published: 10 March 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Marine pollution caused by ship oil spills is a significant issue, accounting for approximately 12% of all pollutants in the ocean. Despite a decrease in ship accidents in recent years due to increased legislation and environmental awareness, oil spills still occur. The impact of these spills on the marine environment highlights the need for further research on the topic [1–3].

Studies on the issue of ship oil spills have mainly focused on the risk of spills and emergency preparedness. However, few have been conducted on the effectiveness of clean-up of the pollutants after a spill. Effective clean-up can significantly reduce the impacts of a spill, protect the environment, and promote sustainable development.

This study aims to identify the factors that impact the effectiveness of clean-up operations after a ship oil spill. The analysis utilizes historical spill accident data from the United States Coast Guard (USCG) database and employs a Tree-Augmented Naïve (TAN) Bayes approach for variable selection and classification. This approach overcomes the subjectivity by using data-driven techniques rather than relying on expert judgement to derive state probabilities.

The results of this study provide guidance for improving the effectiveness of clean-up operation in the event of a ship oil spill. This is a crucial aspect of emergency response and will help to minimize damage to the marine environment.

The rest of this paper is structured as follows. Section 2 provides a review of the relevant literature on ship spills and TAN methods. The methodology used in this study is outlined in Section 3. In Section 4, the model-building process is detailed, including information on the dataset for the TAN model, its design, and an interpretation of the variables. Section 5 presents the results of the model, including sensitivity analysis and evaluation for different train/test ratio and scenario tests. The study's findings and recommendations for future research are summarized in Section 6.

2. Literature Review

2.1. Study on Oil Spills from Ships in the Marine Environment

The risk of ship oil spill is a major threat to the marine environment and can occur during a ship's navigation and operations. There has been a significant amount of research focused on assessing the risk of oil spills from ships and developing emergency response strategies. For example, Schulze et al. [4] analyzed the risk of ship oil spills by considering factors such as the frequency of ship traffic accidents, the frequency of ship oil spill accidents, and the number of ships, while taking into account the impact of weather conditions. Lee et al. [5] evaluated the risk of contamination by examining both the probability of pollution and the time to first impact of pollution (oil and hazardous and noxious substances). Al Shami et al. [6] divided the pollution risk assessment into three pollution indicators and developed a comprehensive hazard score. In addition, many researchers have examined various factors associated with ship oil spills and used different methods to evaluate their effects. Khan et al. [7] combined binary logistic regression and expert judgment to identify influencing factors and analyzed these factors using Bayesian networks.

In terms of emergency response, many studies have been conducted on building emergency preparedness capacity. Most countries have established offshore oil spill emergency response systems led by the government, implemented by professional organizations, and participated in by relevant industries [8]. Since 1997, the U.S., for instance, has a three-tiered oil spill response system that is led by the U.S. Coast Guard and involves various government agencies [9]. In order to prevent the problem of insufficient emergency capacity of a ship oil spill accident, the emergency response capacity of oil spill was evaluated [10]. Novack et al. [11] used an event tree method to analyze ten different oil spill incidents and developed an evaluation model of emergency and preventive measures. Chai et al. [12] proposed a new evaluation method for offshore oil spill emergency response capability and applied the analytic hierarchy process to determine the weight of each factor. Finally, a case study of the model was carried out, and OS-ERC grade was obtained according to the status quo of each influencing factor in the study area.

Another research direction towards the treatment of oil spills is the use of novel methods and materials to prevent or slow the advancement of oil spills and remove them from the sea. Environmentally friendly materials and techniques are being developed [13]. Zafirakou [14] focused on preventing and detecting oil spills, developing methods for treating the oil spills, and removing them from the natural environment, assessing the impact that such accidents cause on the various forms of life. Dhaka et al. [15] introduced various physical remediation techniques for treating marine oil spills and emphasized the importance of proper use of the equipment. Đorđević et al. [16] studied the importance of proper handling and selection of specialized skimmer equipment for accidental pollution with oils.

There are many factors that will affect oil recovery capability or effectiveness. Zhang et al. [17] found that various factors, such as the working environment of the recovery system, the type of oil spill recovered, the thickness of the oil film, water conditions (wind and waves, tides, temperature), and surface debris, can affect the operation of a ship oil spill recovery system. Zhong et al. [18] showed that the fraction of the oil on the water surface and on the shoreline, as well as the amount of oil recovered, were affected by the time of the initial release, the overall duration of the discharge, wind, and recovery actions. El-Gaya et al. [19] studied the effect of disk material and disk surface roughness

on the oil recovery rate and oil recovery efficiency. Etkin et al. [20] sorted the limiting factors affecting mechanical recovery in offshore settings into three general categories: oil properties and behavior; environmental conditions; operational and logistical issues, so as to study the effectiveness of mechanical recovery for large offshore oil spills. However, there is currently a lack of research on how factors related to ship accidents can impact the effectiveness of oil spill clean-up.

2.2. Data-Driven Bayesian Approach

The use of Bayesian methods in marine risk assessment is well-established, with many studies relying on the subjective judgments of experts [21]. However, in recent years, a data-driven Bayesian approach, known as the Tree-Augmented Naïve Bayes (TAN) methods, has gained popularity in the analysis of factors contributing to maritime accidents. For example, Fan et al. [22] used the TAN method to examine human factors in maritime accidents and identified key risk factors such as ship operation, age, navigation segment, and condition. Wang and Yang [23] applied TAN to study shipping accidents involving small vessels on inland rivers and coastal waterways and determined the significant risk factors that impact accident severity.

Cakir et al. [24] used decision tree (DT) and data-driven Bayesian network (BN) methods to predict the severity of oil spills accidents. Kamal et al. [25] employed a data-driven Bayesian method to analyze probabilistic relationships between factors such as ship age, flag, wind speed, visibility, and currents that contribute to accidents.

This paper uses a data-driven Bayesian approach to identify ship oil spill accident-related factors affecting the effectiveness of clean-up efforts. A TAN analysis model was constructed using a comprehensive historical USCG database of oil spill accidents. Relationships between clean-up ratio and influencing factors were explored. The results provide important managerial insights into optimizing management and planning oil spill clean-up efforts following oil spill accidents.

3. Methods

3.1. Tree-Augmented Naïve Bayes (TAN)

The Tree-Augmented Naïve Bayes (TAN) classifier is a novel approach to Bayesian network classification proposed by Friedman et al. [26]. This classifier builds upon plain Bayesian by adding directed connection arcs between attributes with strong dependencies but slightly restricting the connection relationship of each attribute, so that the graph model expressing the dependencies between attributes presents a tree structure. Overall, compared with the plain Bayesian, the TAN structure can make fuller use of the dependencies between the attribute variables, which not only avoids the exponential computation brought by the complex dependency structure but also can acquire better classification results [27].

The TAN classifier is defined by a set of variables $U = \{X_1, X_2, \dots, X_n, C\}$, where $X = \{X_1, X_2, \dots, X_n\}$ is the set of attribute variables, and C represents the class variable. In the TAN tree structure, the class variable serves as the root and has no parent node, while each attribute variable X_i has at most one other attribute variable and the class variable as its parent nodes C , i.e., the attribute variable may have at most two parent nodes. Let $\prod_{i=1}^n X_i$ denote the set of parent nodes of attribute X_i ; then we have $\prod C = \emptyset, C \in \prod_{i=1}^n X_i$, and $|\prod_{i=1}^n X_i| \leq 2$ (where $i = 1, 2, \dots, n$).

The restriction of each node having at most one non-class parent distinguishes TAN from Bayesian networks and provides two key advantages. Firstly, it can reduce the search space, and secondly, it alleviates the problem of estimating probabilities from the training set, as the size of the conditional probability table can grow exponentially with the number of nodes in a Bayesian network.

When using the model for classification, for an instance $x_i = \langle x_{i1}, x_{i2}, \dots, x_{in} \rangle$ of any unknown class, $P(c_j | x_{i1}, x_{i2}, \dots, x_{in})$ is computed according to the Bayesian formula, and selects the class label c^* that maximizes this probability as the instance's class attribution:

$$\begin{aligned} c^* &= \operatorname{argmax}_{c_j \in C} \frac{P(x_{i1}, x_{i2}, \dots, x_{in} | c_j) P(c_j)}{P(x_{i1}, x_{i2}, \dots, x_{in})} \\ &= \operatorname{argmax}_{c_j \in C} P(x_{i1}, x_{i2}, \dots, x_{in} | c_j) P(c_j) \\ &= \operatorname{argmax}_{c_j \in C} P(c_j) \prod_{i=1}^n P\left(x_{ii} | \prod_{i=1}^n x_{ii}\right) \end{aligned} \quad (1)$$

where the set $\prod_{i=1}^n x_{ii}$ is obtained according to the constructed TAN structure.

In the TAN model, each node may have at most one non-class parent node, $\prod_{i=1}^n x_{ii}$ has two forms: $\prod_{i=1}^n x_{ii} = \{c_j\}$, (x_{ii} has no non-class parent node); $\prod_{i=1}^n x_{ii} = \{x_{is}, c_j\}$, (x_{ii} has non-class parent node). The crucial aspect of using the TAN model for classification is to estimate the three sets of probability values $P(c_j)$, $P(x_{ii} | c_j)$, and $P(x_{ii} | c_j, x_{is})$.

3.2. TAN Structure Learning

In Bayesian network (BN) structure learning, there are three main approaches: expert experience, data-driven, and data fusion. Expert experience relies on experts using their experience and expertise to construct BN structures based on personal cause-and-effect relationships. Nevertheless, subjective decisions can result in uncertainty and bias [28]. The data-driven approach involves training models by importing relevant incident data to obtain conditional probabilities of relevant nodes, allowing the creation of BN structures through machine-learning algorithms learning from datasets. Data fusion combines expert knowledge and data machine learning.

This paper employs a data-driven method for BN structure learning and there are several alternative algorithms available, such as TAN, K2 algorithm, and NB [26]. TAN learning constructs a qualitative BN representing RIF interaction dependencies, while the simple Bayesian network algorithm chooses the category with the highest conditional probability for classification. TAN outperforms plain Bayesian while maintaining the computational simplicity and robustness of plain Bayesian, and is more accurate than other data-driven network construction methods [22]. For this reason, TAN learning is adopted in this paper to construct the BN model.

The optimization problem of learning the TAN structure follows the general procedure proposed by Chow and Liu [29], who used conditional mutual information between attributes. The function can be defined as:

$$I_P(X_i, X_j | C) = \sum_{x_{ii}, x_{jj}, c_i} P(x_{ii}, x_{jj}, c) \log \frac{P(x_{ii}, x_{jj} | c_i)}{P(x_{ii} | c_i) P(x_{jj} | c_i)} \quad (2)$$

where I_P represents the conditional mutual information, x_{ii} stands for the i th state of the attribute variable X_i , x_{jj} denotes the i th state of the attribute variable X_j , and c_i represents the i th state of the class variable. To learn the TAN structure, the optimization problem involves finding the tree of defined functions on X_1, X_2, \dots, X_n to maximize the log-likelihood.

3.3. TAN Parameter Learning

After constructing the TAN structure, perform network parameter learning is required [30]. This step is essential to determine the conditional probability distribution of each node and to analyze the causality of ship oil spills. Parametric learning for BN can be performed using expert knowledge or by learning from sample data. However, expert knowledge may be inadequate, inaccurate, and subjective, affecting the accuracy of the network [31]. Therefore, in this study, the EM (Expectation–Maximization) algorithm is used for parametric learning, as it provides superior results in various cases.

The essence of the EM algorithm is to provide an iterative solution for the Maximum Likelihood Estimation (MLE) of the parameters. The EM algorithm consists of two key

steps: the Expectation (E) calculation step and the Maximization (M) calculation step. In general, the incomplete dataset X is comprised of two parts, a complete part (Y) and the missing part (Z) (i.e., $X = (Y, Z)$), and the log-likelihood function of the complete part of the data can be defined as the following Equation (3):

$$L(\theta) = \log g(Y|\theta) \quad (3)$$

where $g(Y|\theta)$ denotes the likelihood function of the whole part of the data Y , and θ is the set of unknown parameters. The conditional density function of X given Y and θ can be defined as Equation (4):

$$k(Y, \theta) = \frac{f(X|\theta)}{g(Y|\theta)} \quad (4)$$

where $f(X|\theta)$ denotes the likelihood function of incomplete data X . According to Equation (4), Equation (3) can be written in the form of Equation (5) as follows:

$$L(\theta) = \log f(\theta) - \log k(Y, \theta) \quad (5)$$

Given the complete part of the observation Y and the parameter θ obtained from the previous iteration, the expected values of both sides of Equation (5) can be obtained using Equation (6), as in Equations (7) and (8). Then, Equation (6) can be written in the form of Equation (9):

$$E\{L(\theta)\} = E\{\log f(\theta) | Y, \theta^{(k)}\} - E\{\log k(Y, \theta) | Y, \theta^{(k)}\} \quad (6)$$

$$Q(\theta | \theta^{(k)}) = E\{\log f(\theta) | Y, \theta^{(k)}\} \quad (7)$$

$$H(\theta | \theta^{(k)}) = E\{\log k(Y, \theta) | Y, \theta^{(k)}\} \quad (8)$$

$$E\{L(\theta)\} = Q(\theta | \theta^{(k)}) - H(\theta | \theta^{(k)}) \quad (9)$$

where $\theta^{(k)}$ denotes the set of parameters obtained from the previous k iterations, and $Q(\theta | \theta^{(k)})$ denotes the Q -function.

According to the results of Jensen's inequality, the M-step of the EM algorithm is to maximize the Q -function as shown in Equation (10):

$$\theta = \operatorname{argmax} Q(\theta | \theta^{(k)}) \quad (10)$$

4. Analysis of Factors Affecting the Effectiveness of Oil Spill Clean-Up

4.1. Data

This research paper analyzes ship oil spills from 2002 to 2015 using the USCG's Marine Information for Safety and Law Enforcement (MISLE) Marine Accident and Pollution Database. This database contains historical inspection data related to marine accident and pollution, collected by the USCG, and offers information on various types of accidents such as grounding, collision, and flooding, types of accident locations, and information related to accident vessels, as well as comprehensive records of marine pollution accidents that primarily occurred in the United States and its territories. It is a valuable source for this paper and has been used in several previous studies [32–36].

The USCG database consists of ten key data files, with four of them being relevant to this study: *MisleVslPoll*, *MisleActivity*, *MisleEvent*, and *MisleVessel*.

First, the researchers considered only those ship oil spill incidents that occurred within 12 nautical miles of the U.S. coastline, excluding accidents in the U.S. inland waterways and those outside the U.S. territorial waters. The pollution incident data were then combined with *MisleVslEvent*, *MisleActivity*, and *MisleVessel*, and the conversion between the files was realized according to the *vessel_id*, *activity_id*, and *case_id* fields contained in each data file.

In order to ensure the accuracy of the data, the researchers deleted incidents with incorrect values for leakage, gross ton, and age of 0, as well as incidents with severe missing records or unspecified information. Ultimately, 874 ship oil spill incidents were included in the final dataset.

4.2. Factors Considered

To assess the effectiveness of emergency response in the event of a ship oil spill, the clean-up ratio, which is the ratio between the removed and the total discharged oil spill amount, is used as the class variable. According to the literature [37,38], ten attribute variables are selected to construct the TAN classifier. These attribute variables are: discharge amount, substance type, accident type, location of the accident, damage severity, season, vessel age, gross ton, vessel type, and hull material. Table 1 provides a detailed description of these variables and the classes they belong to.

Table 1. States of variables in the TAN model.

Variables	Node Name	Status
Clean-up ratio	<i>c_ratio</i>	S1: (0%); S2: (<=50%); S3: (>50%)
Discharge amount	<i>discharge</i>	S1: (<100); S2: (100 to 1000); S3: (>1000)
Substance type	<i>s_type</i>	S1: crude; S2: diesel; S3: fuel oil; S4: gasoline
Accident type	<i>a_type</i>	S1: capsized; S2: collision; S3: fire/explosion; S4: flooding; S5: grounding; S6: material failure; S7: sinking
Location of the accident	<i>location</i>	S1: bay; S2: channel; S3: coastal; S4: gulf; S5: harbor; S6: ocean; S7: strait
Damage severity	<i>severity</i>	S1: undamaged; S2: damaged; S3: total loss
Season	<i>season</i>	S1: spring; S2: summer; S3: autumn; S4: winter
Vessel age	<i>age</i>	S1: (<25); S2: (25 to 34); S3: (>34)
Gross tonnage	<i>g_ton</i>	S1: (<=50); S2: (50 to 200); S3: (>200)
Vessel type	<i>v_type</i>	S1: fishing vessel; S2: freight vessel; S3: passenger vessel; S4: service vessel; S5: tanker
Hull material	<i>h_material</i>	S1: aluminum; S2: fiberglass; S3: steel; S4: wood

(1) Clean-up ratio (class variable) target node in the TAN model

The clean-up ratio refers to the percentage of the total discharged oil being recovered. The higher the clean-up ratio, the less impact it will have on the marine environment, and vice versa. Based on this information, clean-up ratio was calculated and categorized as: 0%, <50%, and >=50%.

(2) Discharge amount

The discharge amount of the oil spill accidents was classified into three categories: 100, between 100 and 1000, and >1000 gallons.

(3) Substance type

The type of oil spill substance refers to the physical and chemical properties of the pollutant. In the study, only refined petroleum products (diesel, fuel oil, gasoline, etc.) and crude oil were considered; non-petroleum products, such as chemical and garbage, were not included. The variables were divided into four categories: *crude*, *diesel*, *fuel oil*, and *gasoline*.

(4) Accident type

The USCG database gives a comprehensive taxonomy of accident types, which includes 26 distinct accident categories. In this study, accidents of similar nature were combined into seven major accident types: *capsized*, *collision*, *fire/explosion*, *flooding*, *grounding*, *material failure*, and *sinking*.

(5) Location of the accident

The location of the accident may affect the effectiveness of clean-up actions. The USCG database records the type of waterway with their names where the accident occurred, which were converted into seven types of locations: *bay, channel, coastal, gulf, harbor, ocean,* and *strait*.

(6) Damage severity

This paper classified the damage severity based on the USCG database into three categories: *undamaged, damaged,* and *total loss*.

(7) Seasons

The USCG database records dates of incidents which is converted into seasons as follows: March 21 to June 21 for *spring*, June 22 to September 22 for *summer*, September 23 to December 22 for *fall*, and December 22 to March 20 for *winter*.

(8) Vessel age

Previous research has demonstrated a negative correlation between the age of a ship and its safety level [39]. The likelihood of a vessel being involved in an accident due to structural breakdown increases with age [40]. We categorized the ship ages based on the USCG database as (<25 ages), (25 to 34), and (>34).

(9) Gross tonnage

Gross tonnage (GT) is the unit of measurement for a ship's internal volume. One hundred cubic feet is used as the unit for GT. According to the Maritime Accident Investigation Branch's (MAIB, 2010) statistics, the total risk of loss for small ships is significantly higher than for large ships. Considering MARPOL, we used GT to define ship size as (<50), (50 to 200), and (>200).

(10) Vessel type

Different ships have different safety records. Knapp and Franses [41] demonstrated that fishing boats, yachts, and sailing boats are the types of boats with higher accident rates [42]. The USCG database provides a comprehensive classification of vessels, with 25 distinct categories. In this paper, the vessel types were re-classified into five categories: *fishing vessel, freight vessel, passenger vessel, service vessel,* and *tanker*.

(11) Hull material

As widely accepted by the industry, hull material types were classified based on the USCG database into four types: *aluminum, fiberglass, steel,* and *wood*.

4.3. Construct TAN

Our TAN model employs a total of 11 variables, including the target node, to demonstrate the relationships between them and the target variables (i.e., clean-up ratio). The TAN networks were created using the learning network function of the Netica software tool (Norsys, <http://www.norsys.com> (accessed on 4 March 2023)), which is a practical and user-friendly software for Bayesian network applications. Netica is a practical software and easy to use especially for BN applications which provides a functional user interface. The networks are used to perform various kinds of inferences utilizing the most modern and fastest algorithms [24]. After training the qualitative structure using sample data, the edges between the nodes were identified and the initial TAN model is shown in Figure 1.

There are six steps to construct the qualitative part of a TAN network. The construct TAN procedure is described as follows:

- (1) Select *ratio* as the class variable and *discharge, season, location, a_type, v_type, h_material, age, g_ton, severity,* and *s_type* as attribute variables.
- (2) Determine the conditional mutual information of all attribute pairs and class variables to define their relationships.
- (3) Construct an undirected graph with attributes and conditional mutual information, as shown in Table 2.

- (4) Consider the edge weights ordered from largest to smallest, construct the spanning tree with the most significant weight, and then select the edge from the largest to the most negligible weight without forming a circle. For each selected edge, if this edge is added to form a circle, it will no longer be selected; on the contrary, edges with weights less than this edge will be selected from the more significant weight to the smaller one. Keep the selected edge, and delete the other edges. The weights of the selected edges are shown in bold in Table 2.
- (5) Select the attribute node with the most giant spanning tree as the target node, and set the direction of all arcs to other attributes outward to convert the undirected tree into a directed tree.
- (6) Construct the TAN model by adding class variables (i.e., *ratio*) to the tree and adding arcs of class variables to each attribute. The proposed initially TAN model is shown in Figure 1.

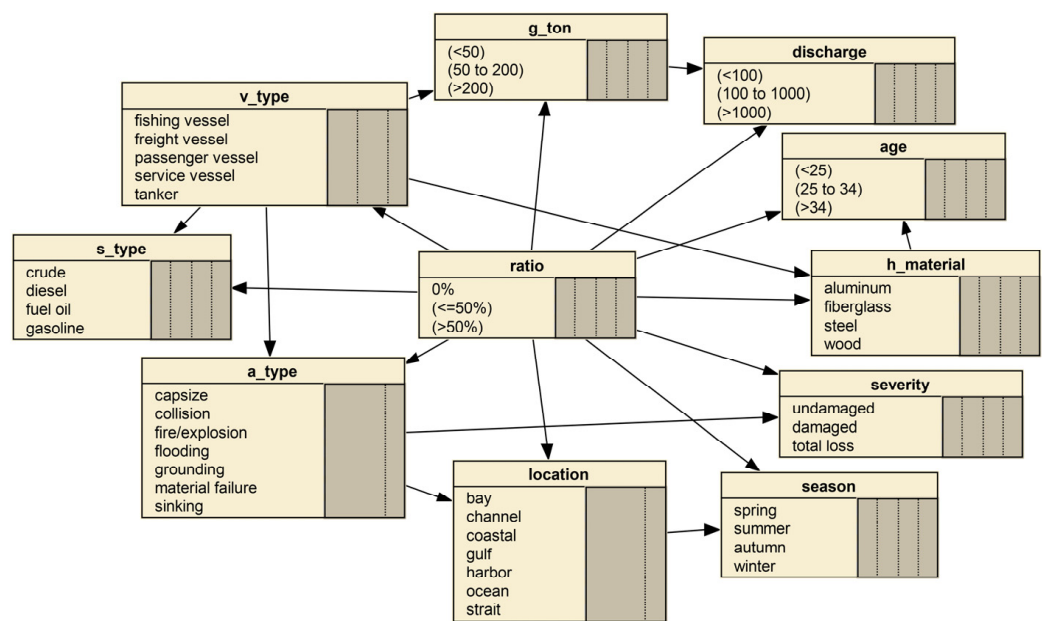


Figure 1. TAN model to assess the factors affecting the effectiveness of oil spill clean-up (i.e., clean-up ratio).

Table 2. Weight of the selected edges to determine the conditional mutual information of the attribute variables.

	<i>v_type</i>	<i>a_type</i>	<i>Discharge</i>	<i>Location</i>	<i>s_type</i>	<i>Severity</i>	<i>Season</i>	<i>g_ton</i>	<i>h_material</i>	<i>Age</i>
<i>v_type</i>	0	0.16	0.09	0.07	0.14	0.07	0.03	0.31	0.33	0.06
<i>a_type</i>		0	0.08	0.10	0.08	0.13	0.03	0.12	0.10	0.07
<i>discharge</i>			0	0.07	0.08	0.06	0.01	0.15	0.12	0.01
<i>location</i>				0	0.03	0.04	0.04	0.05	0.05	0.05
<i>s_type</i>					0	0.03	0.01	0.08	0.05	0.04
<i>severity</i>						0	0.01	0.07	0.04	0.04
<i>season</i>							0	0.02	0.02	0.02
<i>g_ton</i>								0	0.24	0.04
<i>h_material</i>									0	0.11
<i>age</i>										0

5. Results and Discussions

5.1. Model Evaluation

The performance of the model was evaluated using the *F-measure*, *Accuracy*, *Precision*, and *Recall* metrics. The *Overall Accuracy* metric is the most commonly used and efficient

criterion for evaluating a model's prediction accuracy. The accuracy metric evaluates the precision of the classifier's performance, while the recall metric evaluates its consistency. Nonetheless, these metrics are mutually incompatible. To address this issue, the *F-measure*, which balances *accuracy* and *recall* metrics, provides a more accurate evaluation of a model's classification performance. The formulas used to evaluate the model are presented in equations from 11 to 14, with the parameters' meanings listed in Table 3, with evaluation index parameters T_P (True Positive), F_P (False Positive), F_N (False Negative), and T_N (True Negative).

$$\text{overall accuracy} = \frac{T_P + T_N}{T_P + F_P + F_N + T_N} \quad (11)$$

$$\text{precision} = \frac{T_P}{(T_P + F_P)} \quad (12)$$

$$\text{recall} = \frac{T_P}{(T_P + F_N)} \quad (13)$$

$$F - \text{measure} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (14)$$

Table 3. Meaning of evaluation index parameters.

	Predicted Positive	Predicted Negative
Real positive	T_P	F_N
Real negative	F_P	T_N

In order to evaluate the classification accuracy and reliability of the TAN model, it was developed using a training dataset and evaluated with the test dataset. In this study, different portions of the dataset were used to train the model, with the remainder serving as the test dataset for evaluating model performance. Table 4 displays the classification performance of TAN models generated with five different ratios of training and test data. The results of the evaluation indicate that Model 1, which was trained with 90% of the data learned and tested with 10%, has the highest classification accuracy, with an overall accuracy of 0.64. Table 5 provides the performance indicators of Model 1.

Table 4. Performance of TAN models in terms of the accuracy and reliability. The models are constructed in terms of train/test rate.

Model	Train/Test Rate	Accuracy	Ratio	Precision	Recall	F-Measure
Model 1	90/10	0.64	0%	0.56	0.63	0.60
			(<=50%)	0.50	0.11	0.18
			(>50%)	0.69	0.73	0.58
Model 2	85/15	0.58	0%	0.48	0.45	0.47
			(<=50%)	0.33	0.08	0.13
			(>50%)	0.64	0.74	0.56
Model 3	75/25	0.57	0%	0.51	0.47	0.49
			(<=50%)	0.25	0.09	0.13
			(>50%)	0.62	0.72	0.55
Model 4	70/30	0.58	0%	0.53	0.47	0.50
			(<=50%)	0.27	0.12	0.16
			(>50%)	0.62	0.72	0.55
Model 5	65/35	0.58	0%	0.52	0.51	0.52
			(<=50%)	0.20	0.10	0.13
			(>50%)	0.64	0.70	0.56

Table 5. TAN model performance results for Model 1.

Metrics	Clean-Up Ratio		
	0	<=50%	>50%
Precision	0.56	0.5	0.69
Recall	0.63	0.11	0.73
F-measure	0.59	0.18	0.58

The accuracy of 0.64 indicates that the Bayesian model developed in this study correctly predicted the outcome in 64% of cases. Although this accuracy may appear low, it is essential to consider the complexity of the system being modeled and the context in which it is applied. Bayesian models are probabilistic graphical models that can represent and analyze complex relationships between variables based on observed data. Therefore, the accuracy of the model can be constrained by the quality or quantity of available data or the complexity of the relationships between variables.

In this study, we only analyzed ship oil spill accident-related variables, and variables related to emergency response were not considered. This limitation may have impacted the model's predictive performance, which could have been enhanced by incorporating more comprehensive variables and refining the assumptions about their relationships. However, it is important to note that the specific research question of this study was to identify accident-related factors that impact the effectiveness of ship oil spill clean-up, and in this regard, the model performed well, and the modeling approach was appropriate.

Although an accuracy of 0.64 may suggest some limitations in the performance of the model, it is not necessarily a poor result and may be improved with further refinement and validation. The model's predictive performance can be enhanced by using additional data, refining the model structure, and optimizing the model's parameters. Therefore, with further research and analysis, it is possible to improve the model's predictive performance.

5.2. Marginal Probability Distribution of TAN Model 1

With regards to TAN Model 1, which utilized 90% of the data, Figure 2 depicts the marginal probabilities of states for each variable. The probability of more than 50% clean-up ratio was the highest, accounting for 56.2% of all incidents. In the case of a vessel spill, the most common spill amount was between 100 and 1000, accounting for 42.6% of the total events. More than three-quarters (75.1%) of the material type spilled was 'diesel'. A considerable proportion (44.7%) of damage severity was classified as 'damaged'. The most frequent types of accidents were grounding (20.7%), sinking (22.3%), and material failure (22.1%). Fishing vessels had the highest likelihood (52.8%) of being involved in accidents. In terms of vessel size, 51.3% of the boats in the sample are less than 50 GT. Steel (38.9%) and wood (34.5%) accounted for more than 70% of the hull materials. The majority of ship oil spill accidents occurred in channels (26.7%), harbors (25.6%), and bays (26.8%). The largest number of vessels is older than 34 years, reaching 65%. The distribution of accidents across seasons was nearly equal.

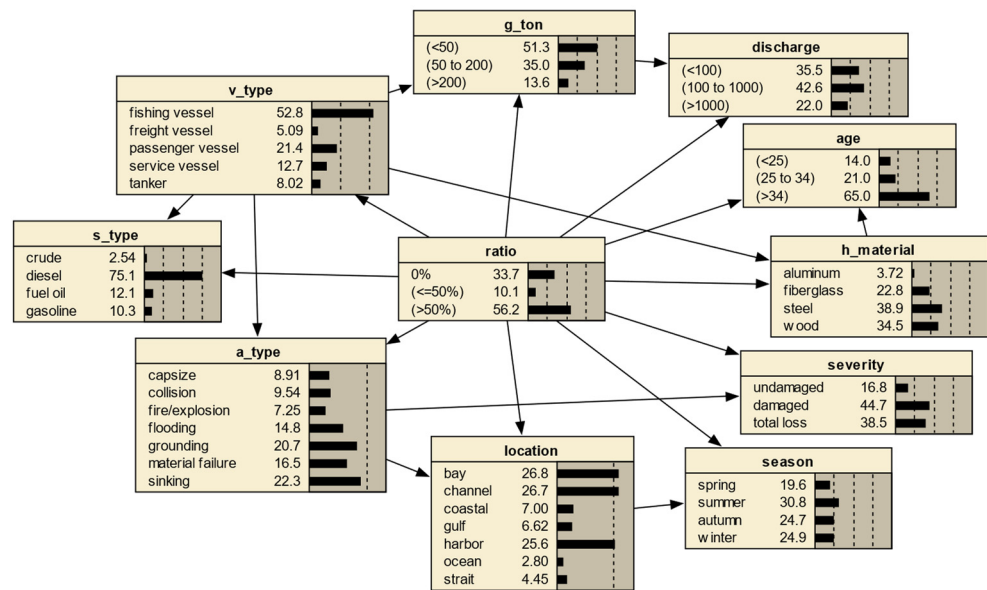


Figure 2. Marginal probabilities for TAN Model 1.

5.3. Sensitivity Analysis

5.3.1. Mutual Information

Sensitivity analysis was performed to analyze the degree of influence of attributes on class variables. Table 6 presents the results of the sensitivity analysis using mutual information. The third column of Table 6 indicates the percentage decrease in entropy. Higher entropy means greater uncertainty in the data. In other words, the nodes with a high potential for reducing entropy potential have a greater influence on determination of the class variable. “discharge” has the most significant impact on the clean-up ratio with an entropy reduction of 4.43%, which can also be interpreted as the target node being the most sensitive to changes in discharge amount. This was followed by “severity” and “location” with entropy reductions of 1.87% and 1.14%, respectively. These results indicate that the clean-up ratio is most sensitive to changes in the state of the discharge amount, damage severity, and location of the accident. “s_type”, “a_type”, “v_type”, and “season” also have a moderate impact on the clean-up ratio with entropy reduction rates of 0.829%, 0.758%, 0.621%, and 0.602%, respectively. However, the effects of “age”, “h_material”, and “g_ton” on the clean-up ratio are relatively small.

Table 6. Sensitivity of the node “ratio”.

Node	Mutual Info	Percent	Variance of Beliefs
ratio	1.32899	100	0.3371677
discharge	0.05881	4.43	0.0138567
severity	0.02487	1.87	0.0066138
location	0.01517	1.14	0.0028081
s_type	0.01102	0.829	0.0022896
a_type	0.01008	0.758	0.0013471
v_type	0.00825	0.621	0.0008535
season	0.00800	0.602	0.0007558
age	0.00656	0.494	0.0002762
h_material	0.00539	0.406	0.0012722
g_ton	0.00279	0.21	0.0002774

5.3.2. Effect of Attribute Variables on Class Variables

The impact of each attribute variable state on the class variables is presented in Figure 3. The first column of each subplot is labeled *average*, which represents the marginal probability of the class variable. The subsequent columns depict the instances

where each state of the attribute variable occurs with a probability of 100%. For example, to assess the effect of the “discharge” state on the clean-up ratio, we can set the ratio of “(<100)”, “(100 to 1000)”, and “(>1000)” to 100% and then compute the new value for each clean-up ratio state. The new values are displayed in the second to fourth columns of Figure 3a. By comparing the second column with the “average” column, if the value of the “0%” class variable in the second column is higher than the “average” column when the status of “(<100)” is fixed at 100%, it can be concluded that the clean-up ratio for vessels that discharge amounts less than 100 gallons may be lower than average.

Figure 3a depicts the impact of spill amount on the clean-up ratio. The graph shows that when the “discharge” state is (<100), the posterior probability of the clean-up ratio is “0%” and then increases gradually from 33.7% to 51.3%. This may be explained by the fact that when the discharge amount is relatively small, the oil film is thin, and the oil layer is relatively sparse, making the collection and removal more challenging. Moreover, emergency response measures may not have been implemented for small-scale spills, resulting in a decrease in the clean-up ratio.

Furthermore, the posterior probability of the clean-up ratio of “>50%” increases from 56.2% to 62.9% and 59.2% for the “discharge” state (100 to 1000) and (>1000), respectively, indicating that within a certain range, the clean-up ratio increased for large spills. This may be because a relatively dense layer of oil is formed when there is a large amount of spilled oil, making it easier to collect and remove. However, if the volume of spilled oil is too high and exceeds the capacity of equipment and personnel, the clean-up ratio may decrease.

Figure 3b reveals that the hull structure in a “total loss” state has a lower probability of a clean-up ratio of “>50%” compared to the “undamaged” and “damaged”. When the damage to the ship’s hull is minor, the amount of spilled oil is relatively small, and the leakage is concentrated, making the clean-up operation easier. However, when the hull damage is severe, the amount of spilled oil is greater and the leakage is more scattered, making clean-up more challenging and possibly leading to lower efficiency. Moreover, the extent of hull damage severity can also affect the difficulty of the clean-up operation. If the hull damage is too severe, there may be a risk to the safety of personnel and equipment involved in the operation.

Figure 3c illustrates that the clean-up ratio is the highest in harbor waterways and the lowest in strait waterways. Additionally, the posterior probability of “0%” clean-up ratio decreases from 33.7% to 25.9% when the accident occurs in harbor waterways. For the waterways of “strait” and “ocean”, the posterior probability of a “0%” clean-up ratio increased to “45.5%” and “45.7%”, respectively, indicating the impact of accident location on the clean-up ratio of oil spills. When an oil spill occurs in a harbor, the clean-up ratio is often higher because personnel and equipment can easily approach and collect the spilled oil. However, if an oil spill occurs at a strait, the clean-up ratio may be limited due to the complexity of the maritime environment and equipment limitations.

Figure 3d displays the impact of pollutant type on the clean-up ratio. The results indicate that lighter petroleum products, such as gasoline and diesel, exhibit higher clean-up rates due to their greater solubility in water and their ease of being washed away by water, making them relatively easy to remove. Conversely, heavier petroleum products, such as fuel and crude oil, have higher viscosity and are less susceptible to water washing, leading to the formation of larger oil films and a greater difficulty in clean-up, resulting in lower clean-up rates.

Figure 3e shows that the clean-up rate for flooding and sinking accidents is higher than for other types of accidents. Flooding and sinking events usually result in a large amount of oil leaking from a specific source, making it easier to identify the oil spill source. On the other hand, the clean-up rate for oil spills resulting from fire and explosion incidents is lower, possibly due to the significant damage to the vessel itself, resulting in oil tank failure and large amounts of oil leaking rapidly. Furthermore, fires and explosions make oil spill clean-up more difficult and dangerous, making effective emergency response and clean-up work more challenging.

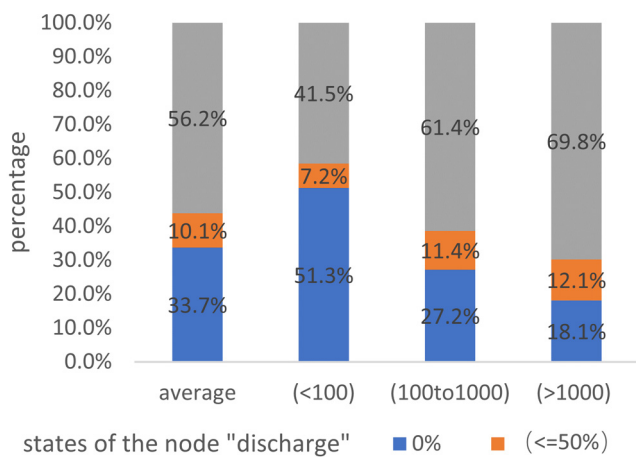
The data in Figure 3f show that oil tankers have a relatively high clean-up rate compared to other types of vessels, while fishing vessels have a relatively low clean-up rate. The higher clean-up rate of oil tankers may be attributed to the fact that they are equipped with professional spill emergency response teams and specialized equipment, including spill response, spill control, and oil pollution cleaning equipment. The crew of oil tankers also receives specialized training in spill emergency response, which enables them to respond quickly to spill incidents and reduce the environmental impact. In contrast, fishing vessels may have limited resources for spill response and may not have the same level of specialized training or equipment, resulting in a lower clean-up rate.

Figure 3g demonstrates that the clean-up rate for oil spills is highest in the summer season. This is likely due to the calm weather conditions and higher temperatures, which facilitate the evaporation and volatilization of the oil pollution, leading to an easier clean-up process.

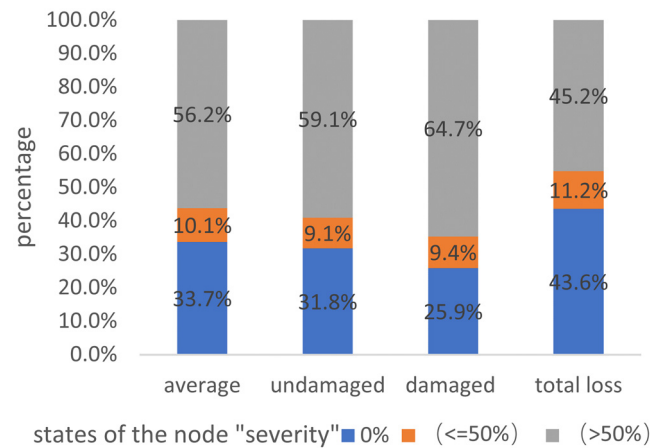
Figure 3h displays that ships in the age group of 25 to 34 have a higher clean-up ratio compared to other age groups. Figure 3i indicates that steel vessels have the lowest clean-up ratio, which may be due to their increased susceptibility to damage and leakage, resulting in larger oil spills that are more challenging to control.

Figure 3j indicates that the clean-up rate of ships with a gross tonnage of less than 50 tons is higher than that of ships over 200 tons. This may be because smaller vessels typically produce smaller oil spills with more concentrated leaks, which are easier to manage and potentially result in higher clean-up rates. Conversely, larger vessels may produce larger oil spills with more dispersed leaks, making clean-up operations more difficult and potentially resulting in lower clean-up rates.

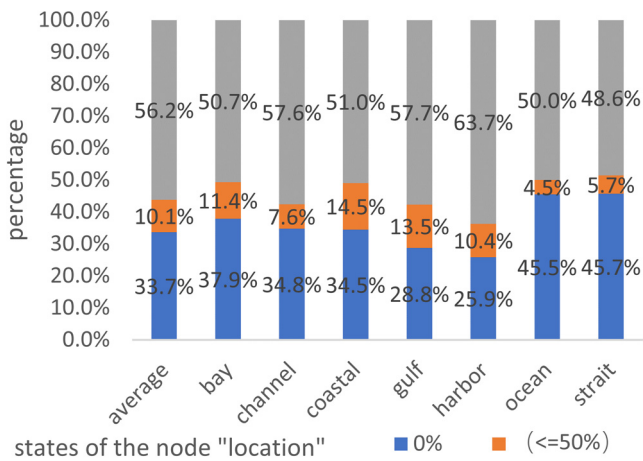
The results of the sensitivity analysis show that the occurrence of a clean-up ratio “>50%”, which has the highest occurrence probability (56.2%), is most strongly associated with vessel hull material being aluminum, a discharge amount of over 1000, and the substance type being gasoline. On the other hand, the occurrence of a clean-up ratio of “0%” (33.7%) is most strongly associated with a discharge amount of less than 100, the location of the accident being in a strait, and the age of the vessel being less than 25. Furthermore, for the clean-up ratio “≤50%” (10.1%), which has the lowest clean-up ratio, freight vessel type, location of the accident being coastal, and activity type being fire/explosion were found to be the most influential factors in that order. These findings suggest that various factors have different impacts on different clean-up ratios and should be considered when developing effective emergency response plans for different types of accidents.



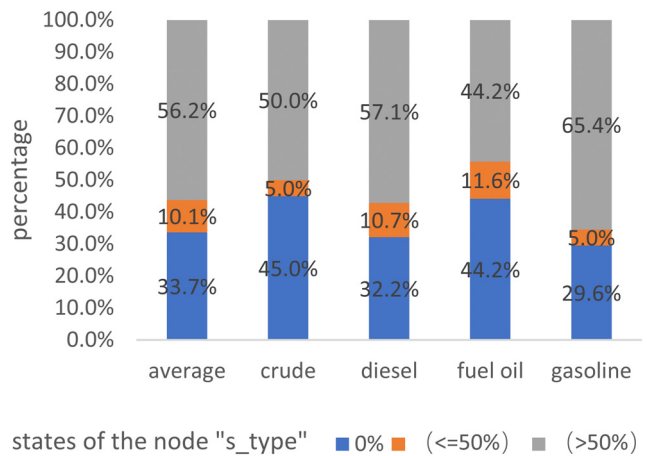
(a)



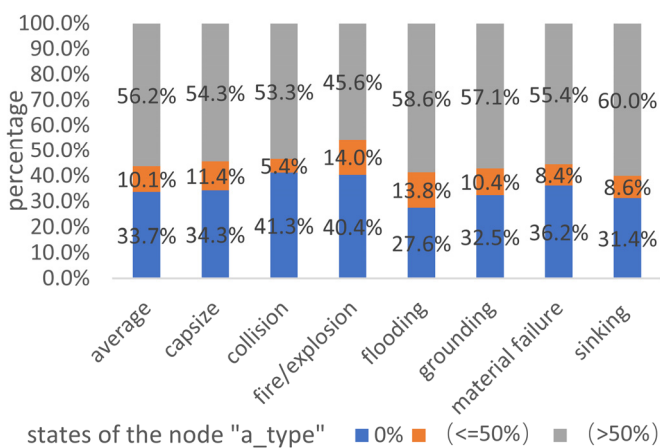
(b)



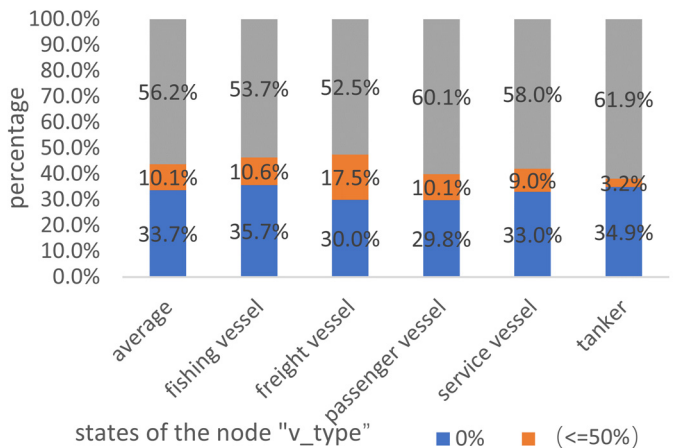
(c)



(d)



(e)



(f)

Figure 3. Cont.

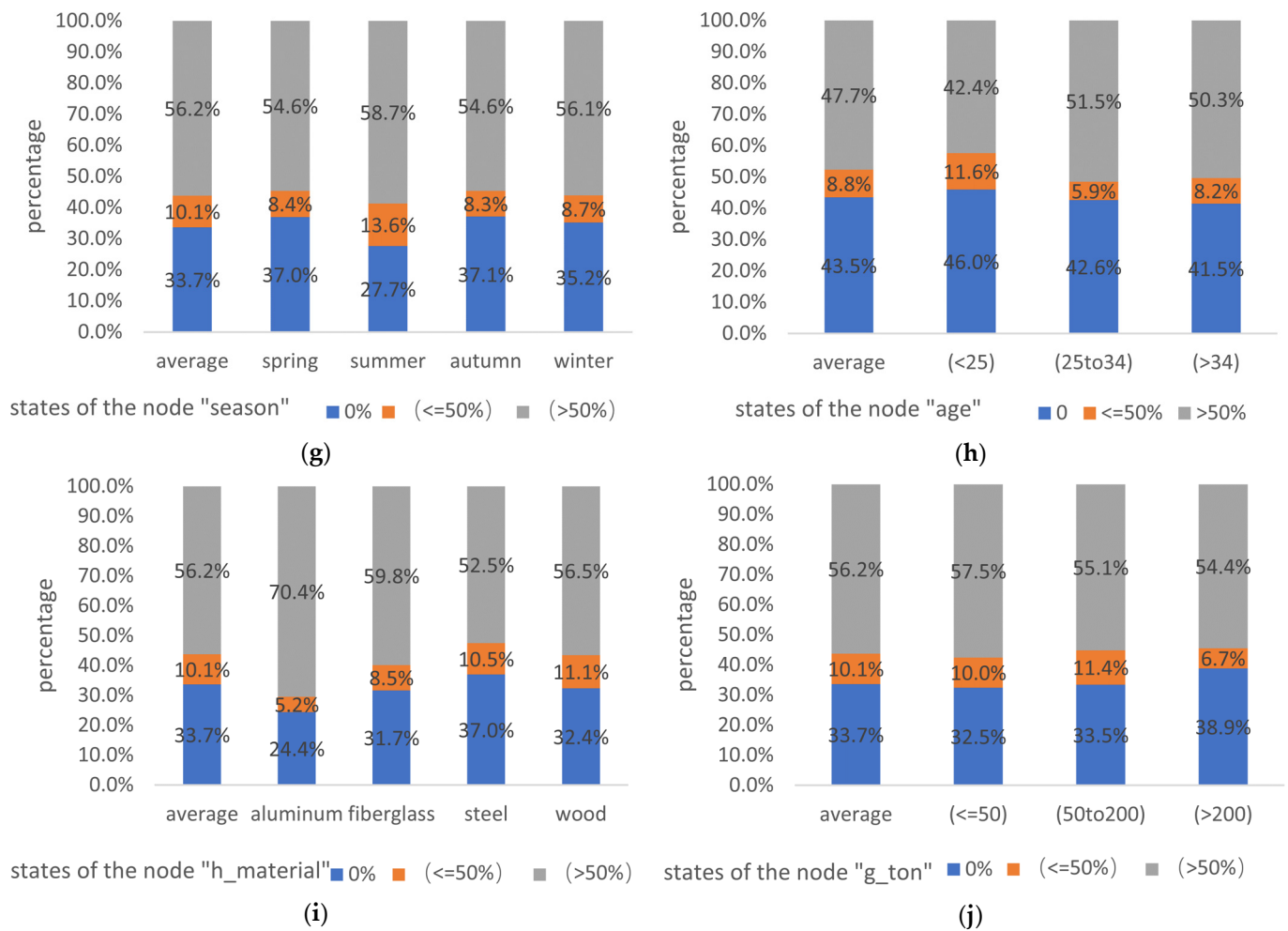


Figure 3. Effect of different states of attribute variables on class variables. (a) The effect of different “discharge” states on “ratio”, (b) the effect of different “severity” states on “ratio”, (c) effect of different “location” states on “ratio”, (d) the effect of different “s_type” states on “ratio”, (e) the effect of different “a_type” states on “ratio”, (f) effect of different “v_type” states on “ratio”, (g) the effect of different “season” states on “ratio”, (h) the effect of different “age” states on “ratio”, (i) the effect of different “h_material” states on “ratio”, and (j) the effect of different “g_ton” states on “ratio”.

5.3.3. Scenario Test

By adjusting the node states used in the Bayesian network (BN) model, our scenario analysis updates the probabilities. This allows us to calculate the probability of different clean-up ratios in various scenarios, which provides us with valuable insights. We analyzed the worst-case scenario, which involves evidence of an accident resulting from an accident type with a discharge amount <100, occurring in a strait. Figure 4 illustrates this scenario. The probability of the clean-up ratio being “>50%” is only 20.7%, while the probability of the clean-up ratio being “0%” is as high as 74.0%. This suggests that the clean-up effectiveness in this case is very low, and it is crucial for relevant clean-up agencies and personnel to take action to increase the clean-up ratio.

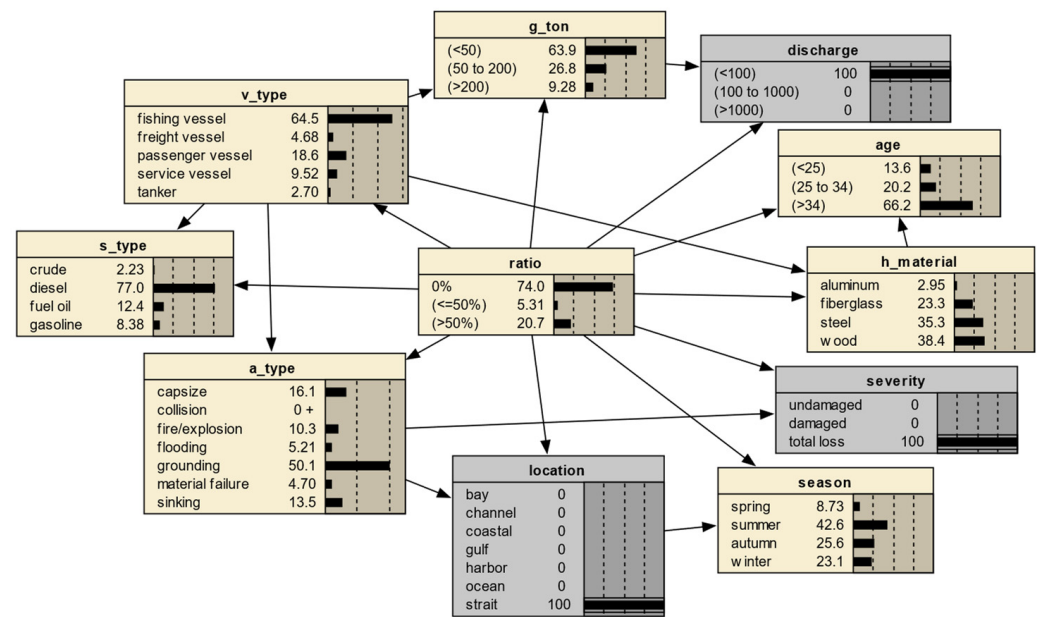


Figure 4. Worst-case scenario derived by combining three nodes: *severity*, *location*, and *amount of discharge*.

6. Final Remarks

The efficiency of clean-up efforts following oil spill accidents from ships can be influenced by several factors that can be broadly categorized into two groups: accident-related factors and emergency response-related factors. Accident-related factors include the scale of pollutant leakage (the amount of oil spilled), accident severity, and ship conditions such as the type of vessel involved and the extent of damage sustained. Environmental factors, such as the location of the accident and weather conditions, also fall under this category. Emergency response-related factors include the availability of clean-up resources, clean-up techniques employed, and the timing of the response. These factors can significantly impact the effectiveness of clean-up efforts and the ability to contain and remove spilled oil.

Previous studies mainly focus on evaluating the emergency response capability following an oil spill accident, and they do not comprehensively examine the accident-related factors that influence the effectiveness although these factors contribute to uncertainty related to the clean-up efficiency. Identifying such factors is a pressing challenge, as it enables the development of more targeted and effective responses to improve the clean-up operation and effectiveness.

To address this gap, our research employs a data-driven Bayesian network (TAN) approach to analyze the pollution clean-up ratio after an oil spill accident and identify the factors that impact the clean-up ratio. Our findings highlight the significance of certain variables in determining the success of oil spill clean-up efforts, such as the discharge amount, damage severity, and accident location. These insights have practical implications for emergency response planning and decision making. This study advances previous works (such as Chai et al. [11], Steven D. Novack et al. [10], and Zhu [13]) by providing new insights into the effectiveness of clean-up operations in response to ship oil spill accidents.

The study has significant implications for sustainability policies and practices related to oil spill clean-up. Oil spills are a significant environmental risk associated with the shipping and transport of oil, and as such, they need to be addressed as part of broader sustainability strategies. By identifying the factors that influence the clean-up performance against ship oil spill accidents, such as the scale of pollutant leakage, accident severity, and location, this study can inform policies and practices aimed at pinpointing these key factors to develop more targeted and effective responses to oil spills, ultimately reducing environmental impacts of ship oil spill accidents.

Moreover, the utilization of a Bayesian network approach in this study has significant implications for sustainability policies and practices. This modeling tool enables the analysis of complex systems and the understanding of the inter-relationships between various variables. Hence, Bayesian networks hold potential for a wide range of sustainability applications, from managing resources to understanding complex systems. Overall, this study emphasizes the necessity for targeted and efficient responses to oil spills resulting from ship accidents, as well as the potential of Bayesian network approaches to inform sustainability policies and practices.

However, the study focuses on identifying factors associated with ship accidents, and it does not account for emergency response-related factors, such as response capacity, and removal techniques. It is essential to acknowledge that the study has some limitations, and future research can enhance the BN model's predictive accuracy by including a broader range of variables and data. Thus, future research should consider comprehensive factors to improve the validity and reliability of the study's outcomes.

Author Contributions: Conceptualization, P.Z., Y.H. and L.Z.; methodology, L.Z., M.G. and B.Y.; software, L.Z., J.W. and Y.W.; analysis, L.Z., J.W., Y.W., Y.H. and M.G.; writing—original draft preparation, L.Z. and J.W.; writing—reviewing and editing, M.G. and P.Z.; supervision, Y.H. and P.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Natural Science Foundation of China (52272334, 61074142), National Key Research and Development Program of China (2017YFE9134700), and EC H2020 Project (690713). We would also like to thank Donghai Academy of Ningbo University for the financial support in publishing this paper.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank to the National “111” Center on Safety and Intelligent Operation of Sea Bridge (D21013), Zhejiang 2011 Collaborative Innovation Center for Port Economy and Ningbo Collaborative Innovation Center for Port and Shipping Services System for the financial support in publishing this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, S.; Grifoll, M.; Estrada, M.; Zheng, P.J.; Feng, H.X. Optimization on Emergency Materials Dispatching Considering the Characteristics of Integrated Emergency Response for Large-Scale Marine Oil Spills. *J. Mar. Sci. Eng.* **2019**, *7*, 214. [[CrossRef](#)]
2. Gong, J.; Pang, L. Application of marine oil spill model based on Monte Carlo method in Laoshan Bay. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *371*, 042003. [[CrossRef](#)]
3. Villalonga, M.M.; Infantes, M.E.; Colls, M.G.; Ridge, M.M. Environmental Management System for the Analysis of Oil Spill Risk Using Probabilistic Simulations. Application at Tarragona Monobuoy. *J. Mar. Sci. Eng.* **2020**, *8*, 277. [[CrossRef](#)]
4. Schulze, R.H. Probability of an oil spill on the St. Marys River. *Int. Oil Spill Conf. Proc.* **1983**, *1983*, 129–132. [[CrossRef](#)]
5. Lee, M.; Jung, J.-Y. Risk assessment and national measure plan for oil and HNS spill accidents near Korea. *Mar. Pollut. Bull.* **2013**, *73*, 339–344. [[CrossRef](#)]
6. Al Shami, A.; Harik, G.; Alameddine, I.; Bruschi, D.; Garcia, D.A.; El-Fadel, M. Risk assessment of oil spills along the Mediterranean coast: A sensitivity analysis of the choice of hazard quantification. *Sci. Total Environ.* **2017**, *574*, 234–245. [[CrossRef](#)]
7. Khan, R.U.; Yin, J.B.; Mustafa, F.S.; Anning, N. Risk assessment for berthing of hazardous cargo vessels using Bayesian networks. *Ocean Coast. Manag.* **2021**, *210*, 105673. [[CrossRef](#)]
8. Murti, C.K. A systems approach to the control of chemical disasters. *Curr. Sci.* **1986**, *55*, 1064–1066.
9. Zhang, Z. The experience and inspiration of the successful establishment of the U.S. offshore oil spill emergency response mechanism. *Pearl River Water Transp.* **2003**, *4*, 22–24. [[CrossRef](#)]
10. Walker, A.H.; Scholz, D.; McPeck, M.; French-McCay, D.; Rowe, J.; Bock, M.; Robinson, H.; Wenning, R. Comparative risk assessment of spill response options for a deepwater oil well blowout: Part III. Stakeholder engagement. *Mar. Pollut. Bull.* **2018**, *133*, 970–983. [[CrossRef](#)]
11. Novack, S.D.; Siu, N.O.; Hill, S.G. The Use of Event Trees in Oil Spill Prevention Applications. In Proceedings of the International Oil Spill Conference, Fort Lauderdale, FL, USA, 1 April 1997; pp. 527–534.

12. Tian, C.; Huang, Z. Assessment of Emergency Response Capability to Oil Spill at Sea. *Navig. China* **2011**, *34*, 98–103.
13. Li, H.; Li, Y.; Yang, W.; Cheng, L.; Tan, J. Needleless melt-electrospinning of biodegradable poly (lactic acid) ultrafine fibers for the removal of oil from water. *Polymers* **2017**, *9*, 3. [[CrossRef](#)]
14. Zafirakou, A.; Themeli, S.; Tsami, E.; Aretoulis, G. Multi-criteria analysis of different approaches to protect the marine and coastal environment from oil spills. *J. Mar. Sci. Eng.* **2018**, *6*, 125. [[CrossRef](#)]
15. Dhaka, A.; Chattopadhyay, P. A review on physical remediation techniques for treatment of marine oil spills. *J. Environ. Manag.* **2021**, *288*, 112428. [[CrossRef](#)]
16. Đorđević, M.; Šabalja, Đ.; Mohović, Đ.; Brčić, D. Optimisation Methodology for Skimmer Device Selection for Removal of the Marine Oil Pollution. *J. Mar. Sci. Eng.* **2022**, *10*, 925. [[CrossRef](#)]
17. Zhang, T.X.; Zhang, D.W.; Ren, L.C.; Xie, W.N. Current Developing Situations of the Vessel Opportunity Skimming System in Domestic and Foreign Countries. *Res. Waterborne Transp.* **2008**, *2*, 10–14.
18. Zhong, X.; Niu, H.; Wu, Y.; Hannah, C.; Li, S.; King, T. A modeling study on the oil spill of M/V Marathassa in Vancouver harbour. *J. Mar. Sci. Eng.* **2018**, *6*, 106. [[CrossRef](#)]
19. El-Gayar, D.; Khodary, M.; Abdel-Aziz, M.; Khalil, M. Effect of disk skimmer material and oil viscosity on oil spill recovery. *Water Air Soil Pollut.* **2021**, *232*, 1–9. [[CrossRef](#)]
20. Etkin, D.S.; Nedwed, T.J. Effectiveness of mechanical recovery for large offshore oil spills. *Mar. Pollut. Bull.* **2021**, *163*, 111848. [[CrossRef](#)]
21. Li, G.R.; Weng, J.X.; Hou, Z.Q. Impact analysis of external factors on human errors using the ARBN method based on small-sample ship collision records. *Ocean Eng.* **2021**, *236*, 109533. [[CrossRef](#)]
22. Fan, S.Q.; Blanco-Davis, E.; Yang, Z.L.; Zhang, J.F.; Yan, X.P. Incorporation of human factors into maritime accident analysis using a data-driven Bayesian network. *Reliab. Eng. Syst. Saf.* **2020**, *203*, 107070. [[CrossRef](#)]
23. Wang, L.; Yang, Z. Bayesian network modelling and analysis of accident severity in waterborne transportation: A case study in China. *Reliab. Eng. Syst. Saf.* **2018**, *180*, 277–289. [[CrossRef](#)]
24. Cakir, E.; Sevgili, C.; Fiskin, R. An analysis of severity of oil spill caused by vessel accidents. *Transp. Res. Part D Transp. Environ.* **2021**, *90*, 102662. [[CrossRef](#)]
25. Kamal, B.; Çakir, E. Data-driven Bayes approach on marine accidents occurring in Istanbul strait. *Appl. Ocean Res.* **2022**, *123*, 103180. [[CrossRef](#)]
26. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian network classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [[CrossRef](#)]
27. Zhang, K. The Research on Tree-Augmented Naive Bayes's Improvement and Its Parallelization. Master's Thesis, Changsha University Science & Technology, Changsha, China, 2018. (In Chinese). Available online: https://kns.cnki.net/kcms2/article/abstract?v=LDpCwvjv3gGpExRW4b3yINSIxLuALHzeYgVd6KytbQPMgeK7XjMM0DqQD4Un4ADCuPdFv2Mn5bhoPUf0CeYBsVCyelGijZ_HTtehurDufpw0ojKmPAG7w==&uniplatform=NZKPT&language=CHS. (accessed on 11 August 2022).
28. Jiang, M.Z.; Lu, J.; Yang, Z.L.; Li, J. Risk analysis of maritime accidents along the main route of the Maritime Silk Road: A Bayesian network approach. *Marit. Policy Manag.* **2020**, *47*, 815–832. [[CrossRef](#)]
29. Chow, C.; Liu, C. Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory* **1968**, *14*, 462–467. [[CrossRef](#)]
30. Zou, X.; Yue, W.L. A Bayesian Network Approach to Causation Analysis of Road Accidents Using Netica. *J. Adv. Transp.* **2017**, *2017*, 2525481. [[CrossRef](#)]
31. Jiang, M.Z.; Lu, J. The analysis of maritime piracy occurred in Southeast Asia by using Bayesian network. *Transp. Res. Part E Logist. Transp. Rev.* **2020**, *139*, 101965. [[CrossRef](#)]
32. Dobbins, J.P.; Abkowitz, M. Use of advanced information technologies for marine accident data analysis visualization and quality control. *J. Transp. Saf. Secur.* **2010**, *2*, 1–13. [[CrossRef](#)]
33. Dorsey, L.T.C.; Wang, B.; Grabowski, M.; Merrick, J.; Harrald, J.R. Self healing databases for predictive risk analytics in safety-critical systems. *J. Loss. Prevent. Proc.* **2020**, *63*, 104014. [[CrossRef](#)]
34. Ketkar, K.; Babu, A. An analysis of oil spills from vessel traffic accidents. *Transp. Res. Part D Transp. Environ.* **1997**, *2*, 35–41. [[CrossRef](#)]
35. Talley, W.K.; Yip, T.L.; Jin, D. Determinants of vessel-accident bunker spills. *Transp. Res. Part D Transp. Environ.* **2012**, *17*, 605–609. [[CrossRef](#)]
36. Yip, T.L.; Talley, W.K.; Jin, D. The effectiveness of double hulls in reducing vessel-accident oil spillage. *Mar. Pollut. Bull.* **2011**, *62*, 2427–2432. [[CrossRef](#)]
37. Lu, L.; Goerlandt, F.; Banda, O.A.V.; Kujala, P.; Höglund, A.; Arneborg, L. A Bayesian Network risk model for assessing oil spill recovery effectiveness in the ice-covered Northern Baltic Sea. *Mar. Pollut. Bull.* **2019**, *139*, 440–458. [[CrossRef](#)]
38. Montewka, J.; Weckström, M.; Kujala, P. A probabilistic model estimating oil spill clean-up costs—A case study for the Gulf of Finland. *Mar. Pollut. Bull.* **2013**, *76*, 61–71. [[CrossRef](#)]
39. Cariou, P.; Mejia, M.Q.; Wolff, F.C. On the effectiveness of port state control inspections. *Transp. Res. Part E Logist. Transp. Rev.* **2008**, *44*, 491–503. [[CrossRef](#)]
40. Yang, Z.S.; Yang, Z.L.; Yin, J.B. Realising advanced risk-based port state control inspection using data-driven Bayesian networks. *Transp. Res. Part A Policy Pract.* **2018**, *110*, 38–56. [[CrossRef](#)]

41. Knapp, S.; Franses, P.H. A global view on port state control: Econometric analysis of the differences across port state control regimes. *Marit. Policy Manag.* **2007**, *34*, 453–482. [[CrossRef](#)]
42. Wang, H.X.; Liu, Z.J.; Wang, X.J.; Graham, T.; Wang, J. An analysis of factors affecting the severity of marine accidents. *Reliab. Eng. Syst. Saf.* **2021**, *210*, 107513. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.