Mazaheri, Arsham; Montewka, Jakub; Kujala, Pentti

Towards an evidence-based probabilistic risk model for ship-grounding accidents

Published in:
Safety Science

DOI:
10.1016/j.ssci.2016.03.002

Published: 01/07/2016

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

Please cite the original version:
Mazaheri, A., Montewka, J., & Kujala, P. (2016). Towards an evidence-based probabilistic risk model for ship-grounding accidents. *Safety Science*, *86*, 195-210. https://doi.org/10.1016/j.ssci.2016.03.002

Safety Science 86 (2016) 195–210

Contents lists available at ScienceDirect

Safety Science

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

Towards an evidence-based probabilistic risk model for ship-grounding accidents

Arsham Mazaheri [a,*], Jakub Montewka [a,b,c], Pentti Kujala [a]

[a] Aalto University, School of Engineering, Department of Applied Mechanics, Espoo, Finland
[b] Finnish Geospatial Research Institute, Department of Navigation and Positioning, 02431 Masala, Finland
[c] Gdynia Maritime University, Faculty of Navigation, Department of Transport and Logistics, 81-225 Gdynia, Poland

ARTICLE INFO

Article history:
Received 4 September 2015
Received in revised form 31 January 2016
Accepted 1 March 2016
Available online 9 March 2016

Keywords:
Ship-grounding
Bayesian Belief Network
Evidenced-based modeling
Strength of knowledge

ABSTRACT

Most of the risk models for ship-grounding accidents do not fully utilize available evidence, since it is based on accident statistics and expert opinions. The major issue with such kinds of models is their limitation in supporting the process of risk-management with respect to grounding accidents, since they do not reflect the reality to the extent required. This paper presents an evidence-based and expert-supported approach to structure a model assessing the probability of ship-grounding accidents, to make it more suitable for risk-management purposes. The approach focuses on using evidential data of ship-grounding accidents extracted from the actual accident and incident reports as well as the judgement elicited from the experts regarding the links and probabilities not supported by the reports. The developed probabilistic model gathers, in a causal fashion, the evidential contributing factors in ship-grounding accidents. The outcome of the model is the probability of a ship-grounding accident given the prior and posterior probabilities of the contributing factors. Moreover, the uncertainties associated with the elements of the model are clearly communicated to the end-user adopting a concept of strength-of-knowledge. The model can be used to suggest proper risk-control-measures to mitigate the risk. By running uncertainty and sensitivity analyses of the model, the areas that need more research for making educated decisions are defined. The model suggests the high-level critical parameters that need proper control measures are complexity of waterways, traffic situations encountered, and off-coursed ships. The critical area that calls for more investigation is the onboard presence of a sea-pilot.

1. Introduction

A model of a system is constructed to help the system operators and decision makers to understand and test the way the system and its components behave in various circumstances. This is particularly important for risk managers as they need to predict the behavior of a system against different controlling measures that are applied to mitigate the involved risks in the system before actually implementing them (IMO, 2002). In this respect, what is truly important for the risk managers as decision makers to know is "how accurately the model can mimic the system's behavior, and how trustworthy the results of the model could be". In this respect, clearly communicating the level of background knowledge (BK) (see Section 4) that is used to build a model and the uncertainty attached to that knowledge is valuable and actually recommended (IMO, 2012; Aven, 2013a; Mazaheri et al., 2014; Montewka et al., 2014b). Evidence-based risk modeling (Fig. 1) is developed to fulfill such need for more realistic models that are based on real-life scenarios and also to communicate the BK that is used to construct the models (Mazaheri et al., 2015b).

Since, unfortunately, the ship-grounding accidents are not so rare in the maritime world, for the review of accidents in the Baltic Sea there is enough evidence at hand to learn from and to use in risk modeling; see for example Kujala et al. (2009) and Sormunen et al. (2015b). Due to the seriousness of its consequences and relatively frequent occurrence, these types of accidents are attracting a lot of attention in the academia, industry, and also among maritime authorities. A number of approaches trying to describe and model ship-grounding accidents are presented in the scientific literature. For the latest review, the reader is referred to Li et al. (2012), Özbaş (2013), and Mazaheri et al. (2014). One finding from reviewing the existing models is that there exists a large variation in the level of usefulness of the models for decision making, having a risk paradigm in mind. This means that there is a need for research on models that mimic

* Corresponding author.
E-mail address: arsham.mazaheri@aalto.fi (A. Mazaheri).

http://dx.doi.org/10.1016/j.ssci.2016.03.002
0925-7535/© 2016 The Authors. Published by Elsevier Ltd.

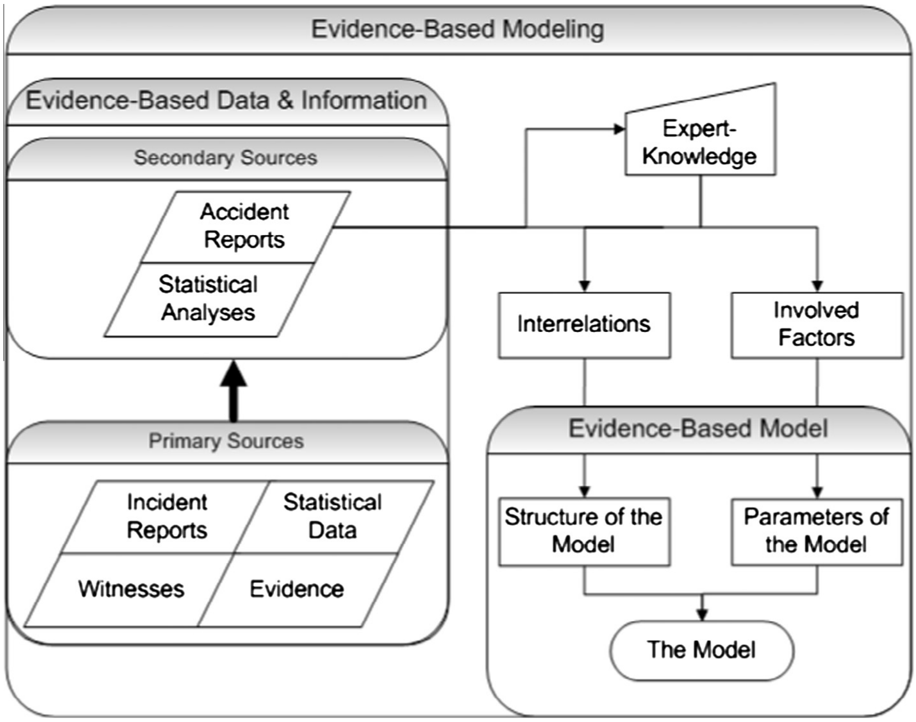

Fig. 1. Framework for evidence-based modeling (replicated from Mazaheri et al. (2015b)).

the relevant elements of a real system in such a way that interventions are possible and their effects on the outcome (i.e. the probability of a grounding of a ship) can be assessed. Second, the use of models that discover causal relations existing in the modeled system and allow two-way reasoning (i.e. from the cause to the effect and vice versa) may be beneficial. Third, the involved uncertainty in all sources of data should be clearly addressed, visualized and presented in the final results. The latter is specifically important as the issue related to the uncertainty assessment is not much discussed in the literature of maritime risk analysis (Goerlandt and Montewka, 2015b; Sormunen et al., 2015a).

As a response to these needs, an evidence-based probabilistic casual model of ship-grounding accidents is proposed here. Contrary to the models that are based purely on expert opinion and thus merely on the intuition of the developers, evidence-based models are supported by real-life scenarios; and the presented BK in such models is assumed stronger (Kristiansen, 2010, Mazaheri et al., 2014). A root cause analysis of the marine accidents in the Arctic region using real accident/incident cases by Kum and Sahin (2015), and quite the same approach for the Shenzhen estuarine waters by Chen et al. (2015) are also recent examples of the response to the need for using real scenarios to mimic the behavior of real systems in modeling. However, an important part of modeling is visualization and communication of the strength of the BK (Goerlandt and Reniers, 2016). This is essential when a risk model is used for decision making and the BK level is weak, thus the uncertainty is high. Such situation may result in risk estimates that falls in the acceptable level, whereas the associated uncertainty may be larger than the margin between risk level and acceptance boundaries, finally resulting in a situation where the risk shall not be deemed acceptable (Montewka et al., 2014b; Aven, 2015). Additionally, the strength of BK is important when assessing the effectiveness of risk control options (RCOs) and the residual risk in the system after implementing the RCOs. Thus, this issue is an additional feature also covered here of the evidence-based modeling.

The objective of this paper is twofold. First, to demonstrate how the collected evidence from the real-life scenarios as accident and incident reports can be used to construct an evidential, probabilistic causal risk model for assessing the probability and contributing factors of such an undesirable phenomena as ship grounding accident. Second, to evaluate the strength of the BK used to develop the model and to communicate it effectively to the end-users to make risk-informed decisions.

To reach the above goal, the remainder of this paper is organized as follow: the grounding model and the methodology that is used to develop it are presented in Section 2. The model validation process is explained in Section 3. A method evaluating the strength of the BK is proposed in Section 4. The model and its results are discussed in Section 5. Section 6 concludes the paper.

2. The model and its development methodology

Considering maritime transportation as a system, a well-defined approach by Haimes (2009) and Aven (2011) can be followed for defining the risk of the system. In that approach, the risk of the system can be defined as $R = \{S, L, C\}$, where S is the scenario for a mishap to occur, L is the likelihood of that specific scenario to occur, and C is the consequence of that specific scenario if it occurs. However, since our knowledge of the system is never complete, the system can never be characterized exactly (Aven and Zio, 2011). Therefore, what we will describe as the risk for a given system, at the end will be formulated merely based on our best knowledge about the system. This incompleteness, which is rooted in our lack of BK on the given system, should always be recognized and communicated. Therefore, the amount of available BK about the system should additionally be considered in the definition of risk. As the result, the description of risk perspective for the given system can yield $R \sim \{S, L, C|BK\}$ (Mazaheri et al., 2014).

Probabilistic causal modeling is known as one of the most suitable methods for modeling the risk of complex systems with high uncertainty like maritime transportation system (Goerlandt and

Montewka, 2015; Hänninen et al., 2014; Montewka et al., 2014; Zhang et al., 2013). Therefore, we adopt here a stepwise approach for constructing a probabilistic causal model, as follows:

1. Define the system with its characteristics and boundaries
2. Collect the required data
3. Develop the model's structure
4. Parameterize the model
5. Validate the model

The approach is proposed by Kragt (2009) and further modified by Akhtar and Utne (2014) to account for the specificity of maritime accidents. Finally, the knowledge extracted from the ship-grounding accident and incident reports (Mazaheri et al., 2015b) is used to develop an evidence-based probabilistic causal model for ship-grounding accidents.

The utilized framework is quite similar to what Montewka et al. (2014a) suggested for risk assessment of maritime transportation systems. In general, the framework suggests to first define what to model and the variables to include, then to develop the qualitative and quantitative parts of the model respectively, and finally to validate the model.

2.1. System definition

The system under the study is marine transportation system. The observable phenomenon within the system is ship-grounding accident that involves the impact of a ship on seabed or on the side of a waterway, due to involvement of various parameters. Therefore, the characteristics and boundaries of the phenomenon under the study can be defined using the superior system as maritime transportation. Since humans are still an important and unavoidable element of such system, maritime transportation is considered as a complex socio-technical system (Woods, 1988). Such system can be defined by four main elements: the environment of the system, which affects the second and third elements of the system, i.e. the current state and the operators, and the fourth element, namely the interface of the system, where the other three elements interact (Fig. 2). A phenomenon like ship-grounding happens within the interface of the system, which thus is affected by the other three elements of the system. Therefore, a number of theoretical parameters from each element of the main system can be listed as factors affecting the grounding accident. Examples of such factors are meteorological and traffic conditions, waterway particulars, human and organizational elements, and the specific particulars of the ship herself. Nevertheless, adoption of parameters that are commonly believed in but which are not based on evidence should be made with caution; since those parameters can be anything presented in support of an assertion and thus would require an uncertainty ranking score. Caution should be applied for the sake of model credibility, its complexity, and the involved uncertainty, both in the model itself and in its outcomes. This issue is further clarified in the following sections as data collection and model construction.

2.2. Data collection

After defining the system, one needs to collect the required data for the modeling. For an evidence-based modeling approach, one of the available sources is the accident reports that are prepared by expert accident investigators (Schröder-Hinrichs et al., 2011). The richness of the accident reports regarding the information about the existing threats as well as the easy accessibility have made the accident reports a desirable source of data and

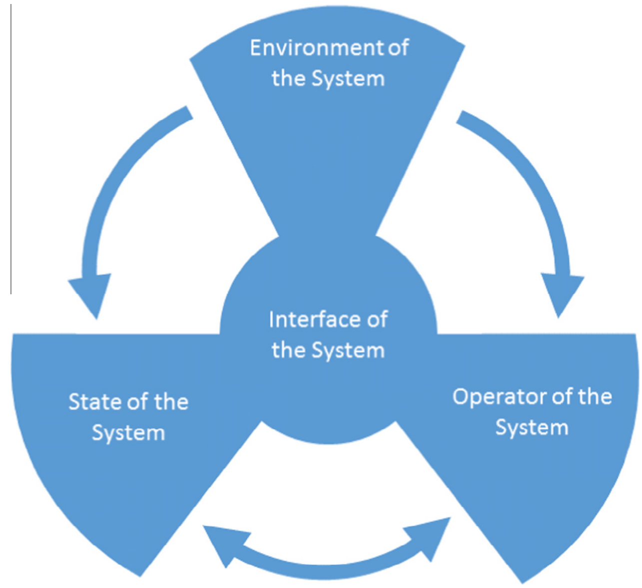

Fig. 2. Maritime transportation as a complex socio-technical system.

information that are based on evidence (Celik and Cebi, 2009, Kristiansen, 2010, Akhtar and Utne, 2014; Mazaheri et al., 2015b). However, despite of all the advantages, using only accident reports for modeling has its own disadvantages. One is that not all the accidents are investigated by the authorities; often only those that may enhance safety awareness among the maritime society are investigated (SIAF, 2014). Therefore, the number of scenarios that can be caught by using accident reports are limited. The other is that, fortunately, accidents are quite rare: this further limits the number of scenarios that can be analyzed (Ladan and Hänninen, 2012). To overcome such downsides, one of the suggested solutions is to utilize incident reports, as incidents[1] take place much more frequently than accidents (Bole et al. 1987). Besides, since incidents are governed by mechanisms and underlying parameters similar to those that cause accidents (Harrald et al. 1998) but do not end up in actual accidents, incidents could be used as an additional source of data which may give insights about the in-placed risk control options that stopped the incident from becoming an accident. Nevertheless, unlike the accidents reports, incident reports are not freely and easily available (Ladan and Hänninen 2012). Moreover, the way that such incidents are investigated affects the quality and thus the reliability of the embedded information.

A recent study by the authors showed that, unlike the mandatory incident reports that are requested and prepared by some companies and authorities, voluntarily incident reports may not be reliable enough to be used for evidence-based risk modeling (Mazaheri et al. 2015b). Therefore, for this study we have only used mandatory incident reports in addition to the accident reports. We have utilized 115 grounding accident reports from the Finnish and British authorities and 90 incident reports from the Finnpilot[2] incident database, which is a mandatory-to-fill database of incidents reported by the Finnish pilots.

[1] Here, an incident refers to an individual mishap or a series of mishaps that did not result in a serious accident like ship-grounding with consequences on human life or the environment, while accident refers to an individual mishap or a series of mishaps that result in an event with minor or major consequences on human life or the environment or causes financial loss.

[2] The Finnish organization responsible for pilotage tasks in Finnish territorial waters.

2.3. Data processing

To develop a grounding model presented here, accident reports are reviewed using the Human Factor Analysis and Classification for Grounding (HFACS-Ground) taxonomy introduced by Mazaheri and Montewka (2014) and Mazaheri et al. (2015b). The incident reports are reviewed using the Safety Factor (SF) taxonomy introduced by Nisula (2014) and Mazaheri et al. (2015b). In the reviewing process of the accident reports, a directed acyclic graph (DAG) is structured for every reviewed accident based on the provided factual information in the report to demonstrate the relation of the contributing parameters in the accident. Therefore, practically for every reviewed accident, in addition to the extracted information using the HFACS-Ground, a DAG is also structured. Finally, all the DAGs are aggregated into a qualitative DAG reflecting the cumulated information from the reviewing of the reports. Due to the shortness and non-comprehensiveness of incident reports, the same procedure is not followed with them. Instead, they are reviewed by using SFs that are high-level positive functions and are believed to be prerequisite for safe transport operations. The SFs are defined with three principles in mind: 1-the factors need to be positive functions and not failure conditions or technical devices; 2-the set should cover all high-level safety–critical functions; and 3-overlap among the SFs should be avoided. In this way, SFs provide an approximation of the real system functions and do not go in-depth compared with other methods like HFACS. Besides, the positive nature of the SFs, as opposed to the failure condition taxonomies used in methods like HFACS, helps to look for the measures that were present in the incident scenarios and presumably stopped the situation to become an accident (Hollnagel, 2015). These features of SFs provide a suitable platform with a proper taxonomy for analyzing incident reports that are not prepared in a systematic analytical way. Therefore, the contributing factors on the incidents as well as the in-placed control measures that stopped the incidents to become accidents are gathered and cumulated using the SFs taxonomy. Thereafter, since the accident and incident reports are reviewed using different approaches, in order to be able to combine the BK, the results are transferred into a common terminology using general categories that are based on failure terminology. Categories from Rothbaum (2000) and McCafferty and Baker (2006) are used as guidelines in building the general categories for this part. In order to decrease the amount of information loss resulting from the accumulation process, the general categories are defined dynamically, i.e. when all the extracted contributing factors from the reports are at hand, they are assigned to pre-created categories. In case that some contributing factors cannot be fit to an existing category or a change to the taxonomy of an existing category is needed in order to accommodate a contributing factor, a new category is created. When a new category is created, all the contributing factors are checked again to see if the new category can be a better fit for any of the previously assigned contributing factors. For a thorough explanation of the data collection and the procedures, the reader is referred to our earlier work reported in Mazaheri et al. (2015b).

2.4. Model construction

The third step in the framework is the setting up a qualitative causal model. In Bayesian Belief Network (BBN) modeling, this step refers to choosing the relevant parameters for the model as nodes as well as the way those nodes are interconnected, that is, the links between the parameters as edges. This was an iterative process. The variables that are extracted as the contributing parameters in the reviewed accident/incident reports in the previous step form the initial nodes of the model. Initially, all the nodes are given two states: the presence or absence of the contributing parameter. This is a simple and safe way to discretize the variables when they are supported by the reports; each report only expresses whether a parameter was present at the time of the accident/incident or whether it was absent. However, during the iteration process and during the parameterization phase (see the next section, "Model Parameterization"), the states of the nodes are adjusted when more information is obtained. For instance, the node "Authority Gradient" (see Fig. 3 and Table A1) was initially given the states "exist" and "not exist"; however, it was later discovered from the reports that the "Authority Gradient" can have three states: "Steep", "Negative" and "Optimal". The first means that the decision of a higher rank officer surpasses all the others' opinions on the bridge. The second refers to situations where the decision of a higher rank officer is always affected by the others' opinions. The third state occurs when other possible opinions on the bridge are also heard and the decision is made by the superior considering the heard opinions. Another example is the node "Meteorological Condition" (see Fig. 3 and Table A1), which initially had two states: "Good" and "Bad". However, when it appeared that this variable can affect the vision of the officer on watch ("Visibility") as well as the navigational equipment ("Signal Quality"), the states were changed to "Storm" that can affect the "Signal Quality", "Fog" that can affect the "Visibility", and "Good" for a normal situation (DNV, 2003; Hänninen et al., 2014). This way of discretization is performed purely based on the evidence at hand, and other sort of discretization that might be regarded as rather arbitrary is avoided at this study. It might be thought that the model can be more accurate if finer discretization (i.e. more states for a variable) or even a continuous distribution were used for some of the variables; nevertheless, one should bear in mind that larger databases with more data are needed if either of the above methods should be utilized to evidentially support finer discretization; a greater number of variable states means more complex and larger conditional probabilities that need to be obtained.

After defining the contributing parameters as the nodes and their states, a link has been made between every two parameters that are statistically dependent. The approach taken to determine statistical dependence is described in the Appendix.

Thereafter, the resulting network is iteratively cross-checked with the individual DAGs of the accidents that are structured in the previous phase in order to define the direction of the added links as well as to retrieve the parameters and connections that may have been lost during the combination process. The combination process is a process in which the individual parameters are accumulated into the global parameters used in HFACS-Ground and SF. As the result, every node and link in the final model is supported either by at least two accident reports or by statistical dependencies; (see Appendix and Table A1). Four nodes and their links (i.e. *Waterway Complexity*, *Location*, *Season*, and *Traffic Distribution*) are supported by the statistical analysis performed earlier by the authors on HELCOM grounding accident data between the years 1989–2010 (Mazaheri et al., 2015a). The *Traffic Distribution* node is additionally supported by the data from the reviewed accident report (see Table A1). The resulting model has 32 nodes (excluding the *Grounding* node) as contributing factors on the grounding accident. The nodes are interconnected with 49 edges, including the links to the *Grounding* node (Fig. 3).

2.5. Model parameterization

After constructing the qualitative model, the fourth step is to quantify the model, that is, fill the prior and posterior probabilities of the nodes.

In principle, there are two common practices to fill the conditional probability tables (CPTs) of a BBN. These practices are not necessarily exclusive, but each has its own uncertainties involved.

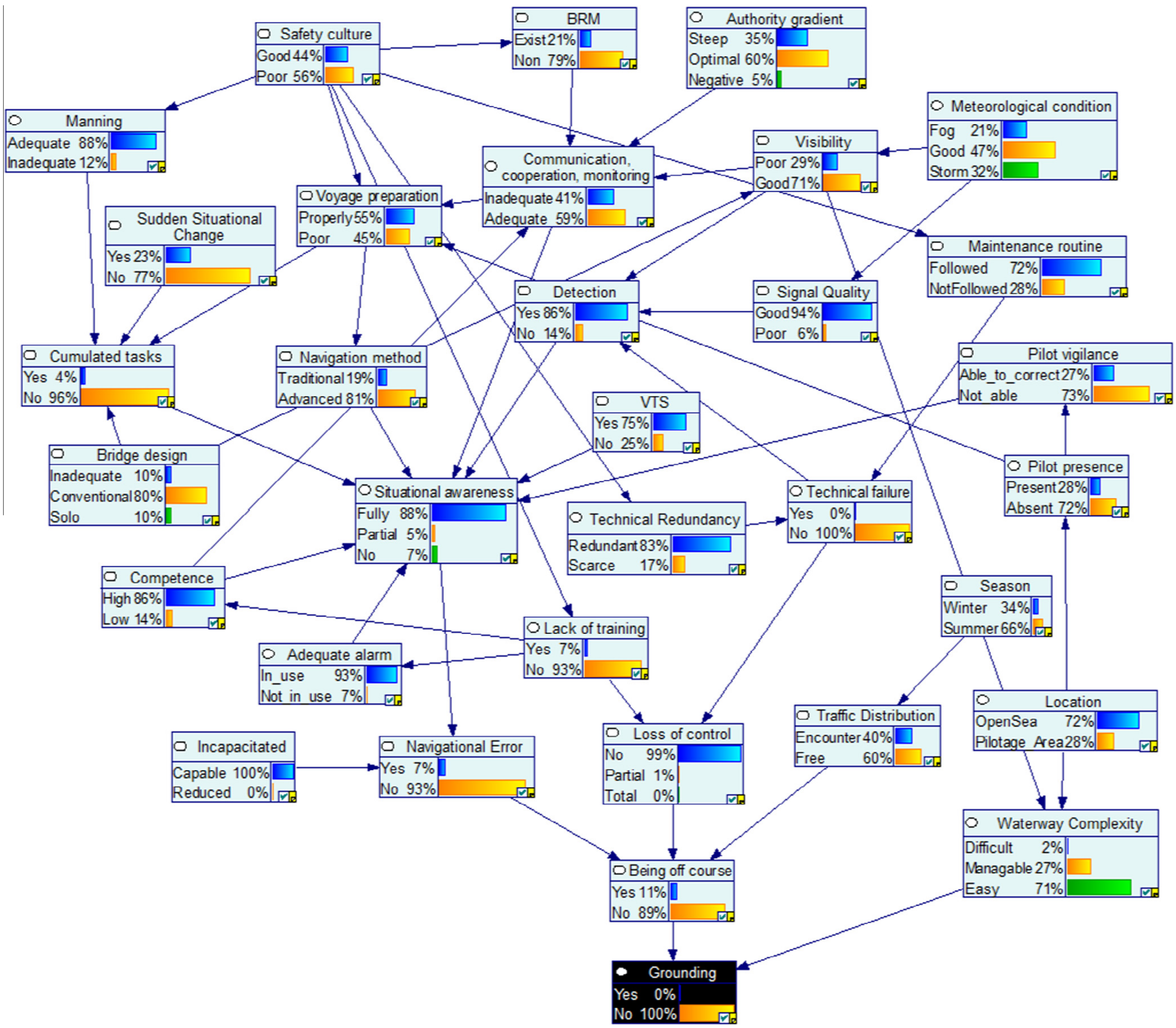

Fig. 3. An evidence-based BBN model for grounding accidents. The probabilities shown as bar charts represent the conditional probability of each node.

One common method is to use techniques that allow elicitation of experts' opinions on the likelihood of the occurrence of each scenario defined by the model, see for example DNV (2003), Cooke and Goossens (2004), O'Hagan et al. (2006), Goossens et al. (2008), and Hänninen et al. (2014). Another way is to use the statistical data and try to find the statistics related to each scenario represented by the constructed model. However, there are several issues that need to be considered when one wishes to use the statistical data. First, there is the fact of accidents' under-reporting. Based on studies regarding the under-reporting phenomena in the maritime realm (Thomas and Skjong, 2009, Psarros et al., 2010; Hassel et al., 2011), one can judge the related uncertainty and thus it can be compensated.

Second, not all the accidents are investigated. According to the Safety Investigation Authority of Finland, not all the reported accidents to the authority are investigated (SIAF, 2014). Therefore, using the accident reports as the statistics for estimating the CPTs can under- or overestimate the probabilities, which translates to uncertainty. Unfortunately, the studies for under-investigated cases do not exist, thus the related uncertainty remains unknown.

Third, each chain of events leading to an accident is different. The ships involved, crew, external and internal conditions, the situational context – in each accident these are different. This means that the exchangeability of events, which is a prerequisite for the frequency assessment, is not valid (Apostolakis, 2014).

Keeping in mind the above discussion and the limitation of the available data on under-reporting and under-investigation and taking into account the lack of exchangeability we decided to utilize the experts' judgment to fill the CPTs of the constructed model.

The CPTs are thus filled using the probabilities presented in the existing literature that have utilized expert opinions for defining the conditional probabilities of different scenarios for a grounding accident (DNV, 2003; Haapasaari et al., 2014; Hänninen et al., 2014). Additionally, the previous studies of the authors regarding the statistics of the grounding accidents (Mazaheri et al., 2015a, b) are used as additional sources. In this regard, if the conditional probability of a scenario in the constructed model of this study is recognized in either of the mentioned sources, the probabilities are filled using the same values presented in the mentioned sources. However, in case that a scenario in the constructed model

Table 1
Type of validity tests for BBN models adapted from Pitchforth and Mengersen (2013).

	Validity type	Description
Construct validity	Nomological	Confident that the model domain fits within a wider domain established by the literature
	Face	Approval from the domain experts that the model looks as expected
	Content	Confident that the available information is integrated into the model via the nodes, links, and discretization as well as parametrization of the nodes
Criterion validity	Concurrent	Confident that the whole network or sections of it behave identically to sections or another network from a similar domain
	Qualitative features test	Showing that the model structure, discretization, and parameters are similar to the nomologically proximal models
	Behavior sensitivity test	Approval that the behavior of the model as well as its sensitiveness to the constructed components is what is expected from the modeled system

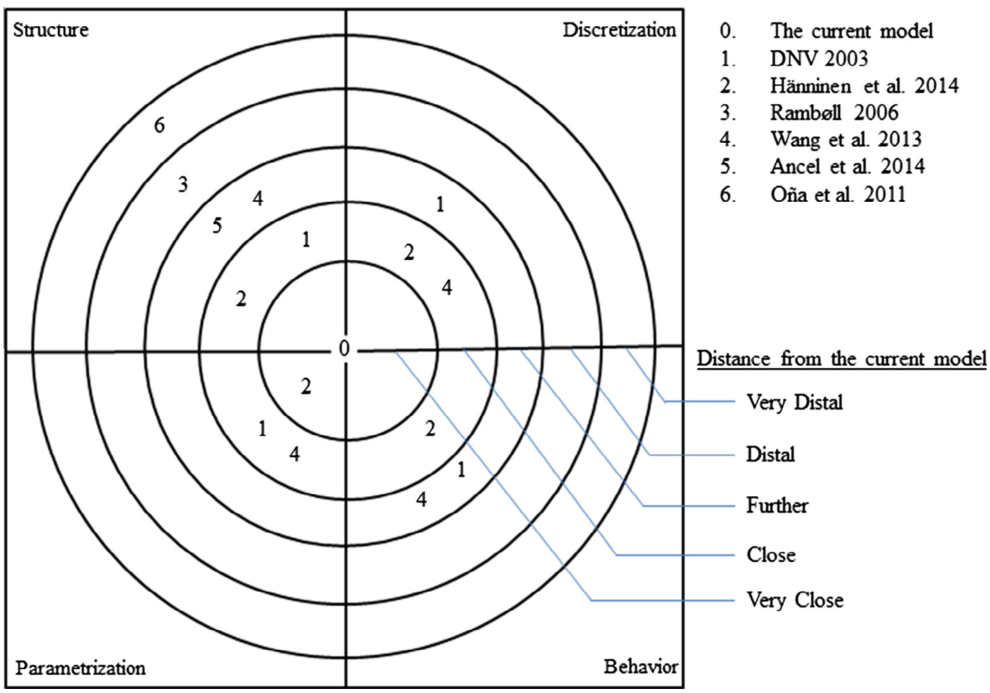

Fig. 4. Nomological map for the BBN model of ship-grounding accident.

is not matched with any scenario presented in the mentioned sources, the probabilities are filled using scenarios similar to those detected in our dataset, which is based on the reviewed accident and incident reports. It means that the probability of the scenario is represented by the number of occurrences of the scenario in the dataset divided by the total number of the reviewed cases (presented in Table A1 in the Appendix as "*205 A&I*"). A factor of 2, supported by the results of Psarros et al. (2010) and Hassel et al. (2011), is utilized to adjust the number of the accident cases to ease the uncertainty effect from under-reporting. Nevertheless, the under-investigated uncertainty is not addressed in estimating the probabilities. As the result, 1065 prior and posterior probabilities are estimated and fed into the model for its quantification.

3. Model validation

Validation of a model means finding whether the built model actually fulfills the purpose for which it was built, given that it confers the specification of the system that it is representing (Marwedel, 2011; Wentworth, 2012).

Model validation can be understood in a wider sense than as a comparison with observed data, by inspecting the model *qua* model. Such approaches to validation are widely used in social science research (Trochim and Donnely, 2008), system dynamics modeling (Forrester and Senge, 1980), expert-based Bayesian Network modeling (Pitchforth and Mengersen, 2013) and recently in the maritime risk modeling (Goerlandt and Montewka, 2014; Goerlandt et al., 2015; Zhang et al., 2015). A fairly similar holistic validity framework with six types of validity tests is adopted here, as presented in Table 1.

First, it is possible to evaluate whether the model adequately operationalizes the construct it intends to measure, i.e. how well it concretizes the object of inquiry for the given purpose. This is evaluated in terms of nomological and construct validities. The latter is broken down into face and content validities. Face validity is a subjective, heuristic interpretation of whether the model is an appropriate operationalization of the construct. Content validity is a more detailed comparison of the elements in the risk model in relation to what is believed to be relevant in the real system.

Second, a number of specific tests can be performed on the model to evaluate whether the model adequately meets certain criteria. Here we refer to criterion validity. A Behavior Sensitivity Test (BST) is used to assess to which model elements the results are sensitive. Parameter sensitivity of a model can be calculated and the results can be evaluated by domain experts. In a qualitative features test (QFT), the model response is evaluated for a number of test conditions in terms of a qualitative understanding of how the system is believed to respond under these conditions. In a concurrent validity test, the model elements are compared with the elements in another model for a similar purpose. This can also include a comparison with the output of such a model if the scope of the applications is the same.

The validity tests do not "prove" that the model results are correct; they only indicate the extent to which the model is a plausible representation of the object of inquiry. The model should be plausible enough to serve as a basis for further reflections, leading to deliberative judgments in the risk analysis (Goerlandt and Montewka, 2015a; Zhang et al., 2015).

3.1. Nomological validity

Nomological validity discusses the position of the constructed model within similar existing models in the context of the studied phenomena, which here is ship-grounding accident. This type of validity needs extensive literature analysis to find and understand the existing models in the realm under the study (Pitchforth et al., 2014). In this way, a nomological map is drawn to estimate the distance of the constructed model from the existing models in four areas of uncertainty in BBN modeling, as follows (Pitchforth et al., 2014):

1- Structure of the model, which includes the nodes that are included in the model as well as the way those nodes are interconnected (i.e. the edges).
2- Discretization, which is about the way the states of each node are defined.
3- Parametrization that defines the conditional probabilities of the nodes and their states.
4- Behavior of the model, which describes the uncertainties related to the output of the model.

The distance for each analyzed pair of models is indicative only, showing the position of the model among other solutions, for four analyzed areas (see Fig. 4). To determine the distance, subjective judgment is used (Goerlandt and Montewka, 2015). An extensive literature analysis (Mazaheri et al., 2014) has detected three practical BBN models currently existing for ship-grounding accidents. The models are listed in Table 2 as no. 1, 2 and 3 together with a few recent BBN models from other proximal realms. The nomological map for the models is shown in Fig. 4. However, due to different modeling choices adopted and availability of the models, not all the pairs can be compared in all uncertainty areas.

Given that the current model and its output as probability of grounding (see Table 3) can be placed within the existing models in the literature and observing the distances that the constructed model in different areas of uncertainty has with the existing models in the maritime domain (Fig. 4), specifically in the grounding accident domain, we are confident that our constructed model is nomologically validated.

Table 2
Existing BBN models for ship-grounding accidents detected in the literature analysis and comparable BBN models in other domains.

No.	Nodes	Edges	Source	Domain
0	33	49	The current model	Transportation/maritime/grounding
1	69	107	DNV (2003)	Transportation/maritime/grounding
2	75	136	Hänninen et al. (2014)	Transportation/maritime/grounding
3	35	48	Rambøll (2006)	Transportation/maritime/grounding
4	16	27	Wang et al. (2013)	Transportation/maritime/collision
5	62	85	Ancel et al. (2015)	Transportation/aviation
6	18	30	Oña et al. (2011)	Transportation/road

Table 3
Output of the proximal grounding BBN models as probability of grounding (P_G) compared with the current model's output.

Source	Current model	DNV (2003)	Hänninen et al. (2014)
Output of the model as P_G	1.62E−5	1.04E−5	3.37E−5

3.2. Construct validity

The construct validity of the model comprises face and content validities described below.

3.2.1. Face validity

Face validity is established through an iteration process, explained in the model construction section, when the preliminary established network is iteratively checked with the individual DAGs from the accident reports. In the iteration process, the added nodes and edges in the network are all iteratively checked and confirmed with the individual DAG of each accident. Any possible disagreements between DAGs of different accidents are assessed and confirmed by the researchers who are experts in the field of ship accident modeling. Therefore, and as the result of the utilized methodology, we are convinced that the involved contributing parameters as well as the connected links of the constructed model are validated; thus the face of the final model is validated too.

3.2.2. Content validity

The content validity checks whether the parameters or relationships that are considered as important by the available literature are included in the model. The contributing parameters in the model and their relations are directly extracted from the real grounding scenarios; thus they are supported by evidence. The discretization of the nodes based here on the extracted information from the reports covers the entire state space of the extracted contributing parameters without any gap. Additionally, the model is parametrized using the available expert knowledge as well as all the available data in hand. Therefore, we are confident that the content validity of the model is established as is the construct validity.

3.3. Criterion validity

The criterion validity presented in this section is broken down into three elements, namely concurrent validity, qualitative feature and behavior sensitivity tests.

3.3.1. Concurrent validity

This type of validity is about the behavior of the whole model or a section of it, in comparison with other models developed for a similar purpose.

The nomological map in Fig. 4 shows that the current model is fairly close to the identified proximal models in the literature within the realm of maritime transportation, especially the ones related to ship grounding accidents. It can also be seen in Fig. 4 that the distance of the current model from proximal models is increasing when we move from the models in maritime transportation system towards the models in other modes of transportation. The only exception is Rambøll (2006) model, which although is from maritime domain, the structure of the model stands by the distance from the current model. This, however, is expected as

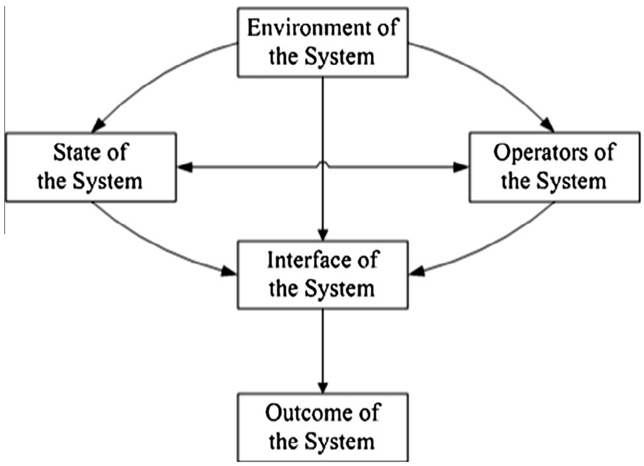

Fig. 5. Components of the current model and their interactions from a system perspective.

although both models are trying to describe same phenomena, the cores of the models are different. The core of Rambøll (2006) model is based on an imaginary geometrical scenarios that are rooted on the scenarios first defined by Pedersen (1995), while the core of the current model is based on the interactions between different parts of the system (see Fig. 5) that themselves are supported by real life scenarios as evidence.

The accident models in the domain normally consist of the sections that account for the factors related to 1-hazardous situation; 2-danger detection; and 3-accident avoidance by evasive actions. The constructed model has 10 general sections as illustrated in Table 4, and all can be fitted within one or two of the groups of factors for similar models in the domain. For instance, sections like ergonomic, traffic, location, and technical failures can all be seen as factors related to hazardous situations; or sections like safety culture, vigilance, and officer on watch are all affecting danger detection; and loss of control, officer on watch, and traffic are sections related to accident avoidance. Sections with similar functionalities also exist in DNV (2003) and Hänninen et al. (2014) models, as the most proximal models in the literature of the same domain. This basically means that all three models follow the same logic and thus should behave the same way. If this observation is combined with the output of the models shown in Table 3, it can assure us that the concurrent validity of our model is acceptable.

Here we assume that the number of nodes associated with each section of the model can represent the complexity of the model.

Table 4
Main sections in the models with the number of the associated nodes for each section.

Section in the model	Current model	DNV (2003)	Hänninen et al. (2014)
Danger detection	5	19	19
Safety culture	6	19	5
Ergonomic	1	2	2
Vigilance	3 (no tug assistance)	7	12
Traffic	1 (traffic distribution)	1 (traffic intensity)	1 (traffic intensity)
Location	1 (waterway complexity)	1 (familiarity of OOW)	1 (familiarity of OOW)
Officer on watch	4	10	12
Technical failures	3	3	4
Meteorological conditions	2	3	3
Loss of control	5	3	6

Table 4 shows that almost all the sections in the constructed model are less complex than those of the compared models, and thus the whole model itself is less complex than those of the compared models. Moreover, the section related to ship traffic in the current model specifically considers traffic distribution, contrary to other two models that consider traffic intensity. This is supported by evidence, as reported in Mazaheri et al. (2015a), that shows the grounding frequency being more dependent on traffic distribution (with respect to a hypothetical central line of a waterway) than on traffic density alone. Additionally, the location sections in DNV (2003) and Hänninen et al. (2014) models only cover the familiarity of the officer-on-watch (OOW) with the waterway, while the same section in the current model covers the waterway's complexity that itself depends on many factors such as draft and speed of the vessel, width of the waterway, number of turns and their magnitude as well as the markings in the waterway (Mazaheri et al., 2015a).

3.3.2. Qualitative feature test

Qualitative features test (QFT) evaluates a model for a number of test conditions to check the model response with the actual understanding of the system behavior under these conditions.

We observed the probability of grounding (i.e. *Grounding* = "*yes*") while setting each state of each variable in a turn. The difference in the probability of grounding for a state producing the largest probability and a state corresponding to the smallest probability is recorded for each variable (Table A2). The difference (ΔP_G) will inform us about the behavior of the model and will help to evaluate the most influential variables:

$$\Delta P_G = \max P(Grounding = \text{“}yes\text{”}|X = x_i) - \min P(Grounding = \text{“}yes\text{”}|X = x_j),$$

where x_i and x_j are the states of variable X that produce the largest and smallest grounding probabilities, respectively. This difference describes the maximum change that variable X could cause on the model output.

The results of this analysis (see Fig. 6 and Table A2) show that there are two variables producing the largest changes in the model output if they are set to either of their states. These are "Being off Course" and "Loss of Control", which seems to follow the expectation regarding the system under analyzed and some other similar systems; see for example Hänninen and Kujala (2012), Hänninen et al. (2012, 2014).

3.3.3. Behavior sensitivity test

Behavior Sensitivity Test (BST) confirms if the model correctly predicts the behavior of the system that is modeled. The test basically checks the sensitivity of the model to see if the model is sensitive to the parameters and scenarios that the system itself is also expected to be sensitive to.

The sensitivity analysis tool of the GeNIe™ software from the Decision Systems Laboratory of University of Pittsburgh is utilized to perform one-way sensitivity analysis, as described in Castillo et al. (1997) and Kjaerulff and van der Gaag (2013). First, the sensitivity function is developed, defining the output probability of interest as a function of the parameter Y; see for example Coupé and van der Gaag (2002); van der Gaag et al. (2007):

$$f(Y) = (c1Y + c2)/(c3Y + c4)$$

where $f(Y)$ is the output probability of interest given observations, and $c1 \ldots 4$ are the constants, which are identified based on the model. Parameter Y is calculated as follows:

$$Y = p(X = x_i|\pi),$$

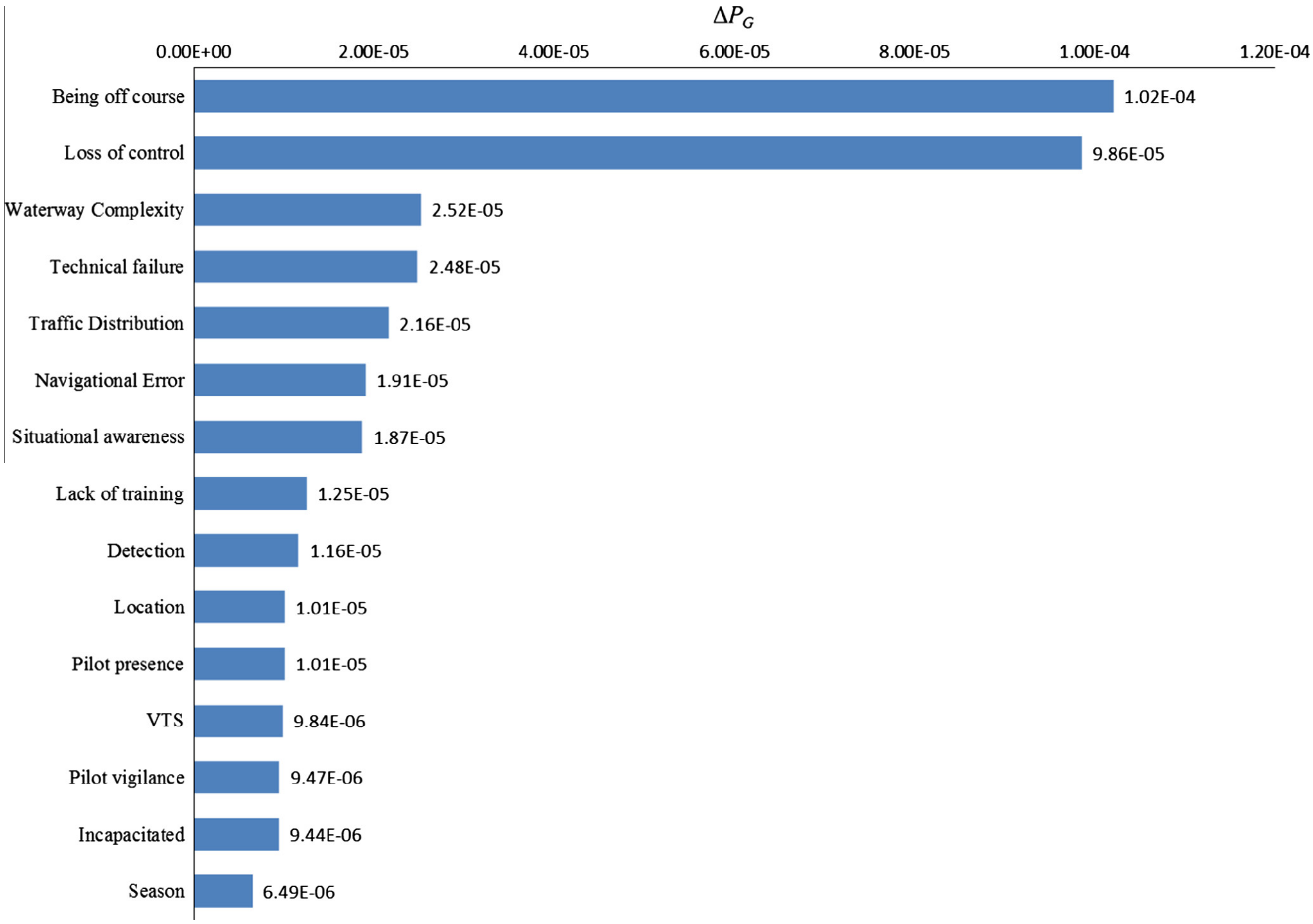

Fig. 6. The top 15 variables producing the largest difference in grounding probability when at their worst state (producing the largest P_G) and best state (producing the smallest P_G).

where x_i is one state of the parameter X, and π is a combination of states for X's parent nodes. When the value of Y is being varied, the other states of the same parameter X; $p(X = x_j|\pi)$, $j \neq i$; should also be varied in order to keep the probability of the sum of the states related to X equal to one (Hänninen and Kujala, 2012).

The first derivative of the sensitivity function $f(Y)$ describes the effect of minor changes in a variable on the output and is called the sensitivity value.

We have analyzed the sensitivity to changes in the posterior probabilities for the following:

- A case where the "Grounding" node is set as a target.
- A case where the "Being off Course" and "Loss of Control" nodes are set as the target nodes. With respect to these two nodes, the model is at its most sensitive, as detected by the previous method.

When the target node is set as "Grounding", there are no evident dominating variables in the model; see Fig. 7. This supports the statement about the complexity of the topic analyzed. On the other hand, the uncertainties associated with the parameters of such complex model affect the model output to a lesser extent. For this reason, such insensitiveness of the model to the changes of the parameters is desirable, as argued by Henrion et al. (2013) and Pradhan et al. (1996).

However, when the target nodes are set to "Being off Course" and "Loss of Control", it becomes evident that the model is sensitive to several nodes, including "Traffic distribution", "Technical failure", "Detection", "Lack of training" or "Incapacitated"; see Fig. 8. Such analysis allows for more detailed assessment of the grounding model and its most sensitive elements.

Evaluating the results of the conducted sensitivity analyses shows that the model sensitiveness to the different parameters of the model and to different combinations of parental states (i.e. scenarios) is in accordance with the expectations of the modeled system. As the result, we are convinced that the final validity test as BST is confidently established so as the criterion validity.

4. The strength of background knowledge

The BK is understood here as a mixture of knowledge, understanding, beliefs and acceptance about the analyzed phenomena; for detailed discussion on this concept the reader is referred to Aven (2013a) and Montewka et al. (2014b). A strong BK will result in a low level of uncertainty, and, intuitively, weak BK causes a high level of uncertainty. In this way, the uncertainty level of the model and its components can be easily communicated to the end-users of the model.

The adopted concept, which originates from the earlier work of Aven (2013b) and is applied by others in the maritime field (Goerlandt and Montewka, 2015a; ValdezBanda et al., 2015), seems the best-suited for our purpose. Assigned to the model parameters, there are three levels of uncertainty based on the amount and quality of the BK that was available as evidence for this particular parameter at the time of the model construction (Table 5).

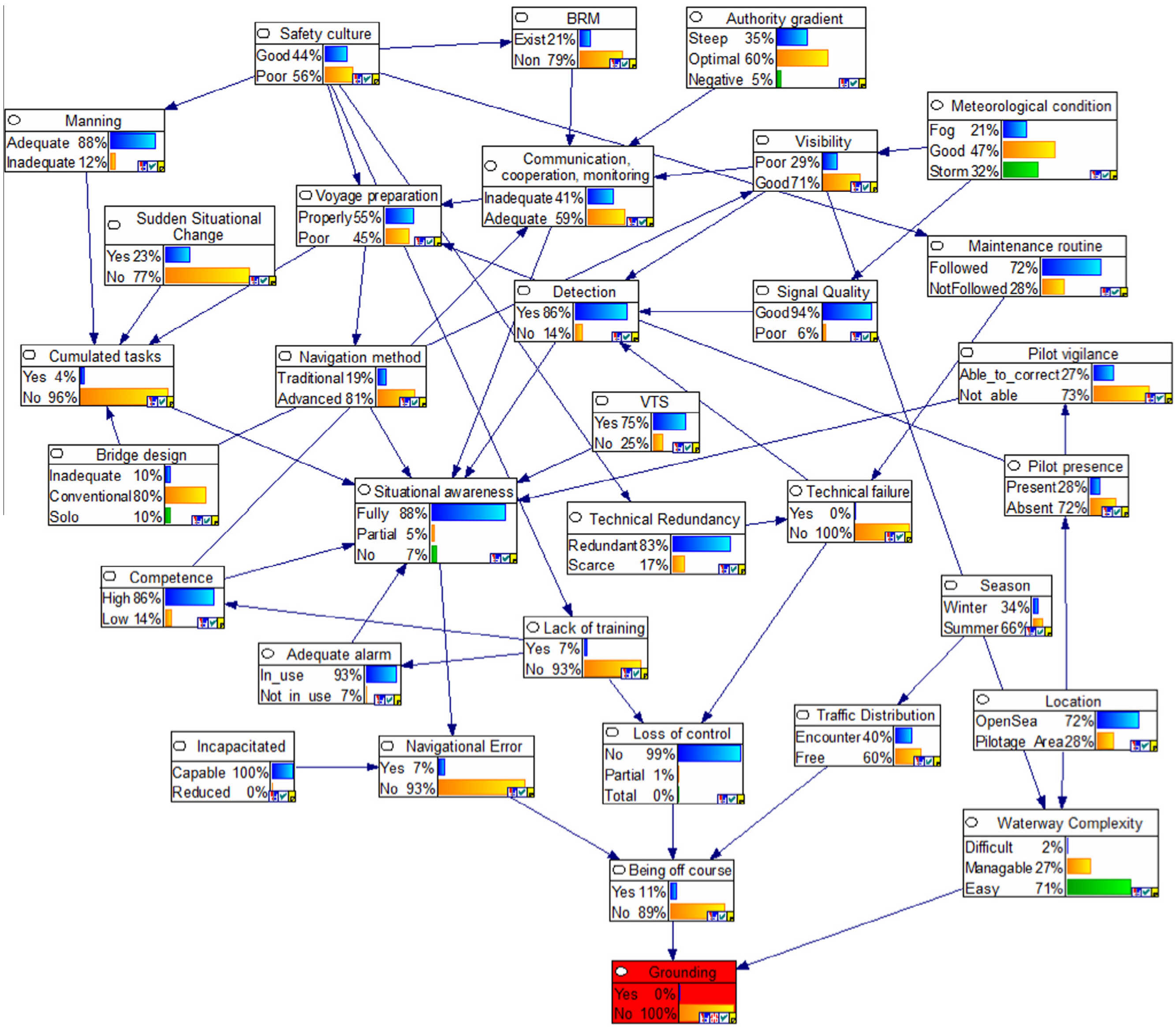

Fig. 7. Sensitivity analysis of the model by setting the grounding as the target node. The model's output as probability of grounding is almost insensitive to all parameters of the model except the Grounding node.

At a simple glance, Fig. 9 visualizes the strength of BK for every parameter (i.e. node) and its parental relations (i.e. edges) of the current model as well as the implemented probabilities based on the sources of evidence that are utilized to support the existence of the parameter or the link in the model (see Table A1). The colors of the border of the nodes show how certain the model makers are regarding the existence of the parameter in the model and the role that the parameter plays in the modeled phenomena. The color of the center of the nodes shows the uncertainty of the implemented prior and posterior probabilities according to the discussion in the model quantification section. Likewise, the colors of the arrows show the certainty of the model makers on the existence of the presented parental relation between the parameters. Therefore, based on the presented strength of the knowledge map of the model (Fig. 9), the outcome of the model can be assessed for the scenarios of interest that are studied as the risk of the system. This basically means that if a certain scenario is being studied, the model can only be used reliably to assess that scenario if the nodes and links that are involved in that scenario have an acceptable level of knowledge strength. Otherwise, the model should not be used for analyzing that specific scenario, or, if it is, the results should be used with caution. The level of acceptability should, however, be decided by the decision maker and according to its utility value (Grabowski et al., 2000).

5. Discussion

In general, it can be argued that the current model stands above the current state of the art. Tables 2 and 4 compare the complexity of the evidence-based model with other existing models. They show that the current model is structurally simpler, with fewer nodes and edges. This makes it easier for the decision makers to understand its mechanism. Moreover, Table 3 shows that the output of the model is not that different from the current beliefs of the domain experts: it has close to the same magnitude of grounding probability as the other two models that are based purely on expert opinions. What makes the model outstanding is its

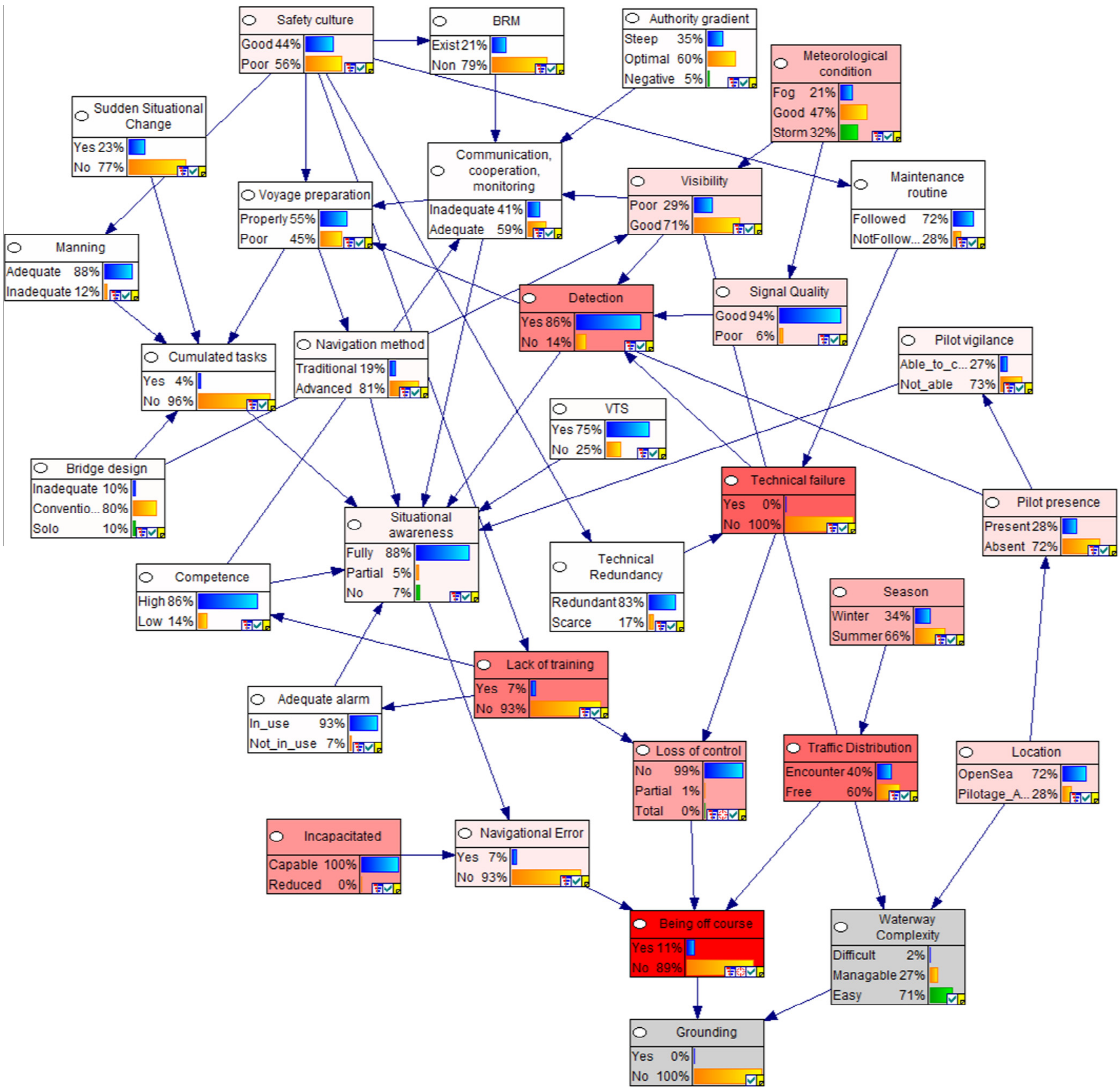

Fig. 8. Sensitivity analysis of the model by setting "Being off Course" and "Loss of Control" as target nodes.

evidence-based feature and two-way reasoning, which allows the evaluation of the effect of high-level interventions that are feasible in real life and their quantification on the model outcome. For instance, the model is able to specify the parameters which affect the outcome the most and estimate the required change in order to reach a predefined level of grounding probability (diagnostic reasoning). Alternatively, the model is able to quantify the grounding probability given a set of input parameters (predictive reasoning). Also, the model encompasses the high-level parameters which are of technical, environmental, personal, and organizational origin. By determining the required change in the probabilities of high-level elements of the model, a more detailed approach can be taken to judge more detailed action. This means that the model could in principle be coupled with other models, e.g. those evaluating the human error or providing the probabilities of technical failure at a lower level of details.

All these are major features of a model that is useful for risk management; that is, the model should be able to suggest and clarify the measures that need to be taken in order to mitigate the risk of accident, in this case ship-grounding (European Transport Safety Council, 2001; Mazaheri et al., 2014). Table 6 is resulted from the developed model and is in line with the above requirement. Table 6 sorts the parameters and their states, based on the magnitude of change in every state of each parameter in order to result in zero probability of grounding. This is obtained by setting the probability of grounding to zero [$P_{(Grounding\ =\ YES)} = 0$] and back-propagating the effect into the model to find the behavior of each parameter in the circumstance.

From Table 6, a decision maker can understand that any risk control option should primarily be focused on developing measures to decrease the likelihood of "going off course" as well as easing the "waterway complexity" and possibility of "traffic

Table 5
The guideline employed for categorizing the strength of the evidence.

Level of knowledge strength	At least two of the following conditions are met
HIGH	High number of supporting cases as evidence (normally more than 9 cases) High reliability of the data/information sources used High detected statistical dependencies High accuracy of the method used for the data analysis High acceptance of the data sources used among the domain experts
MEDIUM	Conditions between the characteristics of the high and low strength knowledge
LOW	Low number of supporting cases as evidence (normally fewer than 4 cases) Low reliability of the data/information sources used Low detected statistical dependencies Low accuracy of the method used for the data analysis Low acceptance of the data sources used among the domain

encounter" (i.e. the top three parameters). Additionally, by looking at Table 6 a decision maker can instantly see where more information is needed in order to make a more educated decision. The criticality of each parameter (last column in Table 6), which cross-checks the importance of the parameter with the uncertainty of that parameter, indicates the parameters that need more study in order to be fully understood. For instance, the exact parameters and interrelations that affect the "waterway complexity" node, as discussed in Mazaheri et al. (2015a), are still to some degree unclear to the maritime society; or, as also discussed in Mazaheri et al. (2015b), the positive effect of "pilot presence" onboard the ships on the probability of grounding can be questioned in some circumstances. Uncertainties as such call for more studies in this regard for discovering the parameters that should be manipulated in order to decrease the likelihood of a grounding accident.

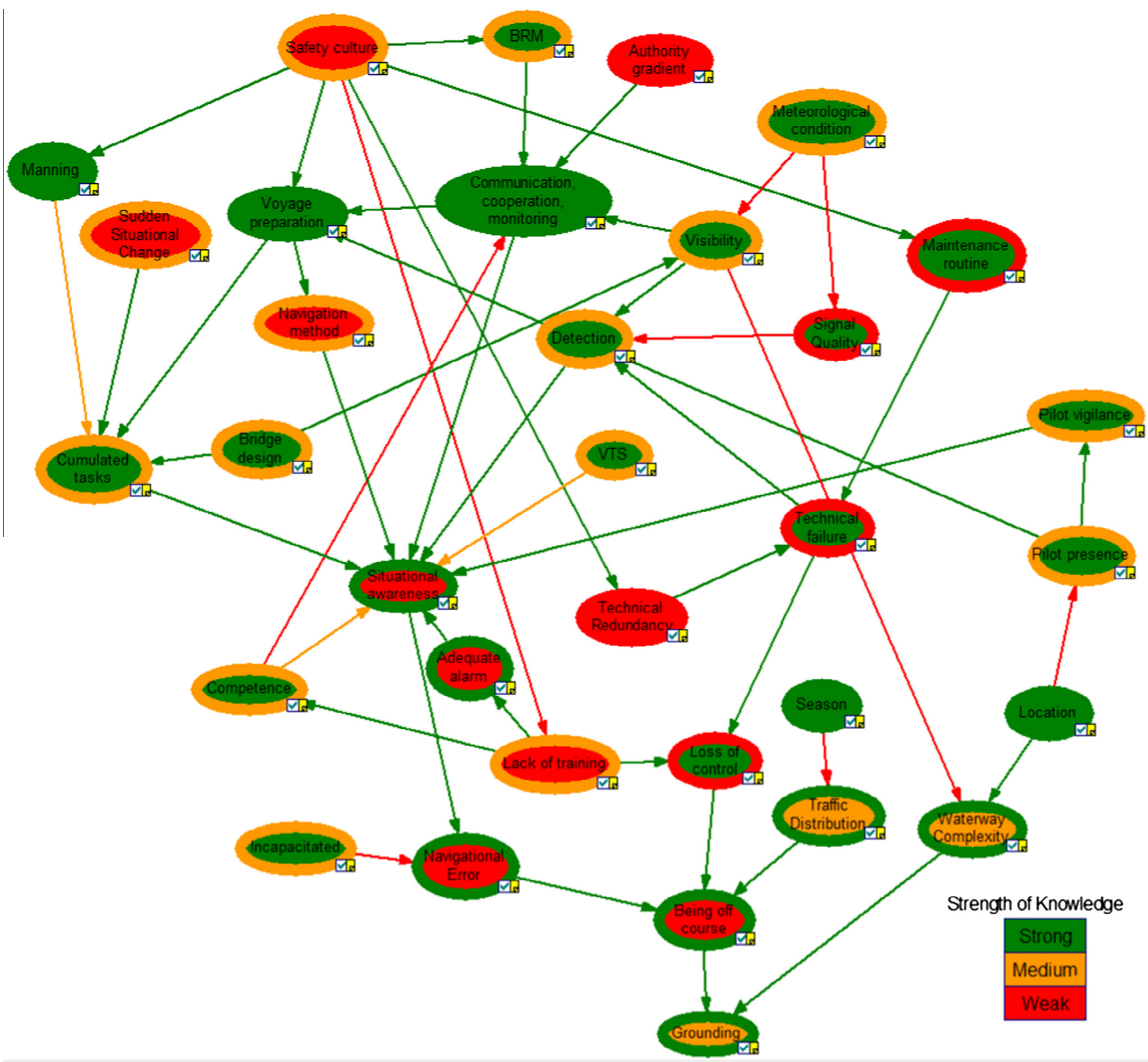

Fig. 9. Strength of the knowledge map of the model (green: strong, orange: medium, red: weak) shows the uncertainty (green: low, orange: medium, red: high) of each node (as the color of the border of the node) and edge (as the color of the link) as well as the conditional probabilities of each node (as the color of the center of the node) based on the sources of the evidence presented in Table 1.

Table 6
Criticality of the parameters of the model when probability of grounding is set to zero [$P_{(Grounding = Yes)} = 0$] with regard to their strength of knowledge and the magnitude of change in their probabilities (S = strong, H = high, M = medium, W = weak, L = low).

Parameter (X)	State (x_i)	δ $P(x_i)$	Strength of Knowledge			Criticality
			Node	Parental	Prob.	
Being off course	<either>	1.00E-05	S	S	W	H
Traffic Distribution	<either>	5.18E-06	S	W	M	H
Waterway Complexity	Easy	2.12E-06	S	M	M	H
Pilot presence	<either>	2.04E-06	M	W	S	H
Location	<either>	2.04E-06	S	-	S	M
Pilot vigilance	<either>	1.85E-06	M	S	S	M
Waterway Complexity	Manageable	1.72E-06	S	M	M	M
Navigational Error	<either>	1.32E-06	S	M	W	M
Situational awareness	No	1.22E-06	S	S	W	M
Situational awareness	Fully	1.16E-06	S	S	W	M
Visibility	<either>	1.02E-06	M	M	S	M
Loss of control	No	8.68E-07	W	S	S	M
Lack of training	<either>	8.21E-07	M	W	W	M
Meteorological condition	Good	7.90E-07	M	-	S	M
Meteorological condition	Fog	6.60E-07	M	-	S	M
Loss of control	Partial	5.23E-07	W	S	S	M
Waterway Complexity	Difficult	4.01E-07	S	M	M	M
Signal Quality	<either>	2.44E-07	W	W	S	M
Safety culture	<either>	1.67E-07	M	-	W	M
Season	<either>	1.46E-06	S	-	S	L
Detection	<either>	1.41E-06	M	S	S	L
Competence	<either>	5.03E-07	M	S	S	L
Loss of control	Total	3.45E-07	W	S	S	L
Voyage preparation	<either>	2.73E-07	S	S	S	L
Communication, cooperation, monitoring	<either>	1.51E-07	S	S	S	L
Meteorological condition	Storm	1.30E-07	M	-	S	L
Navigation method	<either>	9.92E-08	M	S	W	L
BRM	<either>	5.50E-08	M	S	S	L
Situational awareness	Partial	5.41E-08	S	S	W	L
Technical Redundancy	<either>	4.88E-08	W	S	W	L
Adequate alarm	<either>	4.81E-08	S	S	W	L
Maintenance routine	<either>	4.52E-08	W	S	S	L
Manning	<either>	3.76E-08	S	S	S	L
Cumulated tasks	<either>	2.71E-08	M	S	S	L
Authority gradient	Optimal	2.31E-08	W	-	W	L
Authority gradient	Steep	2.06E-08	W	-	W	L
VTS	<either>	2.03E-08	M	-	S	L
Bridge design	Inadequate	1.04E-08	M	-	S	L
Bridge design	Conventional	9.00E-09	M	-	S	L
Authority gradient	Negative	2.52E-09	W	-	W	L
Sudden Situational Change	<either>	2.45E-09	M	-	W	L
Bridge design	Solo	1.41E-09	M	-	S	L
Technical failure	<either>	7.48E-10	W	S	S	L
Incapacitated	<either>	4.72E-10	M	-	S	L

The approach here basically follows the doctrine of utilitarianism, as presented by Kaplan (1997), that helps the decision makers to decide whether they should accept the resulting risk or not. According to Aven (2013b), the strength of knowledge can help the decision makers to decide whether to accept the risk or not only if the calculated risk by the utilized risk model is acceptable for them. Since not all of the decision makers have the same priorities and not even the same degree of feeling about the risk, they have different utility values (Merrick and van Dorp, 2006). The model presented here attempts to help the decision makers to efficiently maximize their own utility values by communicating the level of the BK of the whole model.

Nevertheless, the presented model and its results should not be used without caution. The included parameters do not cover the whole space of the state of the studied system. This is unfortunate but unavoidable as the model is built only based on the evidential data at hand, which beside all issues mentioned earlier is limited temporarily and geographically. Nevertheless, because of the clear communication of the imbedded knowledge into the model, the model has the ability to be developed further by implementing more data in order to cover a larger space of the state of the system. Moreover, the coding system used in the data collection and processing phases was aimed to decrease the subjectivity in the collected data and information as much as possible. However, it should be born in mind that the sources of the data, the accident and incident reports, are still subjective to the understanding of the experts who prepared and reviewed them. This, however, is part of the involved epistemic uncertainty that cannot be eliminated, due to them being inherent in the utilized data sources.

With this being said, what makes the model to stand out from its predecessors is the extensive use of available evidence and clear communication of the involved uncertainty. This makes it possible to define the areas of the model that can be improved whenever more data sources are available. This can be done either by further supporting the current structure with more evidence in order to increase the strength of knowledge behind each parameter and its parental relation, or by modifying the current structure when more reliable and trustworthy sources of evidence suggest adding or eliminating a certain parameter or link.

6. Conclusion

Lack of an evidence-based approach for ship-grounding modeling for building suitable models for risk management purposes was previously shown by Mazaheri et al. (2014) and further supported by Mazaheri et al. (2015a). In this paper, using the actual accident and incident reports of ship-grounding, an evidence-based approach is suggested for building an evidence-based probabilistic causal model for assessing the probability of ship-grounding accidents. The model is believed to be more suitable for risk management purposes as it exposes all the background knowledge behind the construction of the model and thus communicates clearly the involved uncertainties in the model. Each single node and link in the model is supported by actual accident reports, which basically support the integrated scenarios into the model. This has resulted in a simpler structure of the model compared with other existing models of this type.

Besides, the model can be used to suggest areas that need to be controlled by proper risk mitigation measures or areas that need more study to be fully understood. In this regard, the model suggests that the critical parameters that need proper control measures are complexity of waterways, traffic encounters, and a ship being off course. The critical area that calls for more investigation and study, in addition to the above-mentioned parameters, is the onboard presence of a sea-pilot.

Appendix A

A1. Variables existing in the model and their sources

See Table A1.

A2. Statistical evaluation of dependency between nodes in the grounding model

For evaluating the statistical dependency between the nodes in the constructed model, we have assumed a connection between two nodes exists if they show at least 95% significant level of

Table A1
The variables included in the final BBN, their parents, and the source of the evidence for each node and parental relations as edges as well as the prior and posterior probabilities.

No.	Parameter[a] (node)	Parent nodes' no.	Sources (#A = accident reports, #I = incident reports, # the number of occurrences)		
			Probabilities	Node	Parental relations (edges)
1	Adequate alarm	12	205 A&I	15 A	3 A
2	Authority gradient	–	205 A&I	3 A	–
3	Being off course	14, 19, 29	205 A&I	115 A	25 A, Mazaheri et al. (2015b)
4	Bridge design	–	DNV (2003)	6 A	–
5	BRM[b]	22	DNV (2003)	5 A	3 A, Hänninen et al. (2014)
6	Communication, cooperation, monitoring	2, 5, 7, 30	205 A&I, DNV (2003)	15 A	9 A, Mazaheri et al. (2015b)
7	Competence	12	DNV (2003)	6 A	1 A
8	Cumulated tasks	4, 16, 26, 31	DNV (2003)	9 A	18 A
9	Detection	24, 27, 30	Hänninen et al. (2014)	7 A	10 A, Mazaheri et al. (2015b)
10	Grounding	3, 33	Mazaheri et al. (2015a)	–	115 A
11	Incapacitated	–	DNV (2003)	5 A	–
12	Lack of training	22	205 A&I	5 A	3 A
13	Location	–	Hänninen et al. (2014)	Mazaheri et al. (2015a)	–
14	Loss of control	12, 27	DNV (2003)	3 A	6 A, Mazaheri et al. (2015b)
15	Maintenance routine	22	Hänninen et al. (2014)	2 A	1 A
16	Manning	22	205 A&I, DNV (2003)	11 A	3 A, Mazaheri et al. (2015b)
17	Meteorological condition	–	DNV (2003)	5 A	–
18	Navigational method	31	205 A&I	9 A	4 A
19	Navigational error	11, 25	205 A&I	12 A	11 A, Mazaheri et al. (2015b)
20	Pilot presence	13	Hänninen et al. (2014)	5 A	Haapasaari et al. (2014)
21	Pilot vigilance	20	DNV (2003)	5 A	6 A
22	Safety culture	–	205 A&I	9 A	–
23	Season	–	Haapasaari et al. (2014)	Mazaheri et al. (2015a)	–
24	Signal quality	17	Hänninen et al. (2014)	4 A	DNV (2003)
25	Situational awareness	1, 6, 7, 8, 9, 18, 21, 32	205 A&I	20 A	54 A, Mazaheri et al. (2015b)
26	Sudden situational change	–	205 A&I	5 A	–
27	Technical failure	15, 28	Hänninen et al. (2014)	4 A	1 A, Mazaheri et al. (2015b)
28	Technical redundancy	22	205 A&I	2 A	Hänninen et al. (2014)
29	Traffic distribution	23	Mazaheri et al. (2015a)	2 A & Mazaheri et al. (2015a)	Mazaheri et al. (2015a)
30	Visibility	4, 17	DNV (2003)	7 A	3 A
31	Voyage preparation	6, 20, 22	Hänninen et al. (2014)	10 A	1 A, Mazaheri et al. (2015b)
32	VTS[c]	–	Hänninen et al. (2014)	8 A	–
33	Waterway complexity	13, 30	Mazaheri et al. (2015a)	Mazaheri et al. (2015a)	Mazaheri et al. (2015a)

[a] The used terminology is explained in the Appendix of Mazaheri et al. (2015b).
[b] Bridge resource management.
[c] Vessel traffic service.

Table A2
The states producing the largest and smallest grounding probabilities, the corresponding probability values, and their differences for all the parameters of the network.

Parameter X	Best state	P_G^{Best}	Worst state	P_G^{Worst}	ΔP_G
Being off course	No	5.00E−06	Yes	1.07E−E−04	1.02E−04
Loss of control	No	1.54E−05	Yes	1.14E−04	9.86E−05
Waterway complexity	Easy	1.33E−05	Difficult	3.84E−05	2.52E−05
Traffic distribution	Free	1.62E−05	Encounter	4.11E−05	2.48E−05
Navigational error	No	7.63E−06	Yes	2.93E−05	2.16E−05
Situational awareness	Fully	1.48E−05	No	3.39E−05	1.91E−05
Lack of training	No	1.49E−05	Yes	3.36E−05	1.87E−05
Detection	Yes	1.54E−05	No	2.79E−05	1.25E−05
Location	Open sea	1.46E−05	Pilotage area	2.62E−05	1.16E−05
Pilot presence	Present	1.34E−05	Absent	2.35E−05	1.01E−05
Pilot vigilance	Able to correct	1.34E−05	Not able	2.35E−05	1.01E−05
Incapacitated	Capable	1.62E−05	Reduced	2.60E−05	9.84E−06
Season	Winter	1.37E−05	Summer	2.32E−05	9.47E−06
Visibility	Good	1.62E−05	Poor	2.57E−05	9.44E−06
Meteorological condition	Good	1.20E−05	Fog	1.84E−05	6.49E−06
Competence	High	1.48E−05	Low	1.97E−05	4.95E−06
Signal quality	Good	1.46E−05	Poor	1.94E−05	4.80E−06
Voyage preparation	Properly	1.57E−05	Poor	1.98E−05	4.12E−06
Cumulated tasks	No	1.60E−05	Yes	2.01E−05	4.10E−06
Adequate alarm	In use	1.57E−05	Not in use	1.68E−05	1.10E−06
Safety culture	Good	1.62E−05	Poor	1.70E−05	7.41E−07
Navigational method	Advanced	1.56E−05	Traditional	1.63E−05	7.39E−07
Communication, cooperation, monitoring	Adequate	1.59E−05	Inadequate	1.65E−05	6.77E−07
Manning	Adequate	1.61E−05	Inadequate	1.68E−05	6.44E−07
Technical failure	No	1.60E−05	Yes	1.66E−05	6.26E−07
Technical redundancy	Redundant	1.62E−05	Scarce	1.66E−05	3.64E−07
BRM	Exist	1.62E−05	None	1.65E−05	3.40E−07
Maintenance routine	Followed	1.60E−05	Not followed	1.63E−05	3.33E−07
Bridge design	Solo	1.62E−05	Inadequate	1.64E−05	2.23E−07
VTS	Yes	1.62E−05	No	1.63E−05	1.18E−07
Authority gradient	Optimal	1.62E−05	Steep	1.63E−05	9.75E−08
Sudden situational change	No	1.62E−05	Yes	1.62E−05	1.40E−08

Pearson correlation (i.e. P-value < 0.05), following the logic proposed by Akhtar and Utne (2014). Additionally, since Pearson coefficient can only catch the linear dependency between two variables (Lehman et al., 2005), the Mutual Information (MI) test is used to catch the non-linear dependencies between the variables (Steuer et al., 2002; Peng et al., 2005).

In this regard, using MI test, we have calculated the Uncertainty Coefficient (UC) between every two variables. UC determines how large a proportion of the uncertainty about one variable can be decreased by observing the other variable (Theil, 1970), which basically expresses how much two variables are dependent. For this study, every two variables that have UC of more than 10% are considered dependent, and thus their nodes in the model are connected via a link.

Setting the threshold of 10% for UC in order to add a link into the network is justified by knowing that the maximum calculated UC between the parameters of the model was 29%, with the mean value of 2% and standard deviation of 4%. Thus, the 10% threshold is considered as justifiable because it only catches the top 5% of calculated UCs located in the right tail of the distribution of the calculated UCs. This has ensured us that the links with high level of uncertainties are avoided in the model.

A3. Results of sensitivity analysis of the model

See Table A2.

References

Akhtar, M.J., Utne, I.B., 2014. Human fatigue's effect on the risk of maritime groundings – a Bayesian Network modeling approach. Saf. Sci. 62, 427–440.

Ancel, E., Shihb, A.T., Jonesb, S.M., Reveleyc, M.S., Luxhøjd, J.T., Evanse, J.K., 2015. Predictive safety analytics: inferring aviation accident shaping factors and causation. J. Risk Res. 18 (4), 428–451.

Apostolakis, G., 2014. Beware of the Assumptions: Decision Making and Statistics. Probabilistic Safety Assessment and Management – PSAM12, Honolulu, Hawaii, USA.

Aven, T., 2011. On some recent definitions and analysis frameworks for risk, vulnerability, and resilience. Risk Anal. 31 (4), 515–522.

Aven, T., 2013a. A conceptual framework for linking risk and the elements of the data-information-knowledge-wisdom (DIKW) hierarchy. Reliability Eng. Syst. Saf. 111, 30–36.

Aven, T., 2013b. Practical implications of the new risk perspectives. Reliability Eng. Syst. Saf. 115, 136–145.

Aven, T., 2015. Implications of black swans to the foundations and practice of risk assessment and management. Reliability Eng. Syst. Saf. 134, 83–91.

Aven, T., Zio, E., 2011. Some considerations on the treatment of uncertainties in risk assessment for practical decision making. Reliability Eng. Syst. Saf. 96, 64–74.

Bole, A.G., Dineley, W., Nicholls, C.E., 1987. The Navigation Control Manual. William Heinemann Ltd., London.

Castillo, E., Gutierrez, J.M., Hadi, A.S., 1997. Sensitivity analysis in discrete Bayesian Networks. IEEE Trans. Syst. Man Cybern. Part A Syst. Hum. 27 (4), 412–423.

Celik, M., Cebi, S., 2009. Analytical HFACS for investigating human errors in shipping accidents. Accident Anal. Prevention 41, 66–75.

Chen, P., Mou, J., Li, Y., 2015. Risk analysis of maritime accidents in an estuary: a case study of Shenzhen waters. Sci. J. Maritime University Szczecin 42 (114), 54–62.

Cooke, R.M., Goossens, L.H., 2004. Expert judgement elicitation for risk assessments of critical infrastructures. J. Risk Res. 7 (6), 643–656.

Coupé, V.M.H., van der Gaag, L.C., 2002. Properties of sensitivity analysis of Bayesian belief networks. Ann. Math. Artificial Intelligence 36 (4), 323–356.

DNV, 2003. FSA Navigation Large Passenger Ships.

European Transport Safety Council, 2001. Transport safety performance indicators. Brussels.

Forrester, J.W., Senge, P.M., 1980. Tests for building confidence in system dynamics models. TIMS Stud Manage Sci 14, 209–228.

Goerlandt, F., Montewka, J., 2014. A probabilistic model for accidental cargo oil outflow from product tankers in a ship-ship collision. Mar. Pollut. Bull. 79 (1–2), 130–144.

Goerlandt, F., Montewka, J., 2015a. A framework for risk analysis of maritime transportation systems: a case study for oil spill from tankers in a ship–ship collision. Saf. Sci. 76, 42–66.

Goerlandt, F., Montewka, J., 2015b. Maritime transportation risk analysis: review and analysis in light of some foundational issues. Reliability Eng. Syst. Saf. 138, 115–134.

Goerlandt, F., Reniers, G., 2016. On the assessment of uncertainty in risk diagrams. Saf. Sci. 84, 67–77.

Goerlandt, F., Montewka, J., Kuzmin, V., Kujala, P., 2015. A risk-informed ship collision alert system: framework and application. Saf. Sci. 77, 182–204.

Goossens, L.H., Cooke, R.M., Hale, A.R., Rodić-Wiersma, L., 2008. Fifteen years of expert judgement at TU-Delft. Saf. Sci. 46 (2), 234–244.

Grabowski, M., Merrick, J.R.W., Harrald, J.R., Mazzuchi, T.A., van Drop, J.R., 2000. Risk modeling in distributed, large-scale systems. IEEE Trans. Syst., Man, Cybernetics-Part A: Syst. Hum. 30 (6).
Haapasaari, P., Dahlbo, K., Aps, R., Brunila, O.-P., Fransas, A., Goerlandt, F., Hänninen, M., Jönsson, A., Laurila-Pant, M., Lehikoinen, A., Mazaheri, A., Montewka, J., Nieminen, E., Nygren, P., Salokorpi, M., Tabri, K., Viertola, J., 2014. Minimizing risks of maritime oil transport by holistic safety strategies (MIMIC). Final Report. Kotka Maritime Research Center, Kotka.
Haimes, Y.Y., 2009. On the complex definition of risk: a system-based approach. Risk Anal. 29 (12), 1647–1654.
Hänninen, M., Kujala, P., 2012. Influences of variables on ship collision probability in a Bayesian belief network model. Reliability Eng. Syst. Saf. 102, 27–40.
Hänninen, M., Mazaheri, A., Kujala, P., Laaksonen, P., Salmiovirta, M., 2012. The effects of an enhanced navigation support information service on maritime traffic risks in the Gulf of Finland. In: 11th Probabilistic Safety Assessment and Management Conference (PSAM) and the Annual European Safety and Reliability Conference (ESREL), Espoo, Finland.
Hänninen, M., Mazaheri, A., Kujala, P., Montewka, J., Laaksonen, P., Salmiovirta, M., Klang, M., 2014. Expert elicitation of a navigation service implementation effects on ship groundings and collisions in the Gulf of Finland. Proc. Inst. Mech. Eng., Part O, J. Risk Reliability 228, 19–28.
Harrald, J.R., Mazzuchi, T.A., Spahn, J., van Dorp, R., Merrick, J., Shrestha, S., Grabowski, M., 1998. Using system simulation to model the impact of human error in a maritime system. Saf. Sci. 30 (1–2), 235–247.
Hassel, M., Asbjørnslett, B., Hole, L., 2011. Underreporting of maritime accidents to vessel accident databases. Accident Anal. Prevention 43 (6), 2053–2063.
Henrion, M., Pradhan, M., Favero, B.D., Huang, K., Provan, G., O'rorke, P., 2013. Why is diagnosis using belief networks insensitive to imprecision in probabilities. In: Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence, Citeseer, pp. 307–314.
IMO, 2002. Guidelines for formal safety assessment (FSA) for use in the IMO rule-making process. In: Organization, I.M. (Ed.), MSC/Circ. 1023. London.
IMO, 2012. Formal safety assessment, outcome of MSC 90. In: Organization, I.M. (Ed.), Draft Revised FSA Guidelines and Darft HEAP Guidelines.
Kaplan, S., 1997. The words of risk analysis. Risk Anal. 17 (4), 407–417.
Kjaerulff, U., van der Gaag, L.C., 2013. Making sensitivity analysis computationally efficient. In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI 2000), Stanford, CA, USA, pp. 317–325.
Kragt, M.E., 2009. A beginners guide to Bayesian Network modelling for integrated catchment management. Landscape Logic.
Kristiansen, S., 2010. A BBN approach for analysis of maritime accident scenarios. In: Proceedings of the ESREL, Rhodes, Greece.
Kujala, P., Hänninen, M., Arola, T., Ylitalo, J., 2009. Analysis of the marine traffic safety in the Gulf of Finland. Reliability Eng. Syst. Saf. 94 (8), 1349–1357.
Kum, S., Sahin, B., 2015. A root cause analysis for arctic marine accidents from 1993 to 2011. Saf. Sci. 74, 206–220.
Ladan, M., Hänninen, M., 2012. Data sources for quantitative marine traffic accident modeling. In: Aalto University Publication Series SCIENCE + TECHNOLOGY 11/2012. Aalto University, Espoo, p. 68.
Lehman, A., O'rourke, N., Hatcher, L., Stepanski, E.J., 2005. JMP for Basic Univariate and Multivariate Statistics : A step-by-step guide. SAS Press, Cary, NC, USA.
Li, S., Meng, Q., Qu, X., 2012. An overview of maritime waterway quantitative risk assessment models. Risk Anal. 32 (3), 496–512.
Marwedel, P., 2011. Evaluation and validation. In: Dutt, N.D., Marwedel, P., Martin, G. (Eds.), Embedded System Design, second ed. Springer, Dortmund, pp. 203–234.
Mazaheri, A., Montewka, J., 2014. Usability of accident and incident reports for evidence-based risk modeling of ship grounding. In: Proceedings of the Annual European Safety and Reliability Conference (ESREL), Wroclaw, Poland.
Mazaheri, A., Montewka, J., Kotilainen, P., Sormunen, O.-V.E., Kujala, P., 2015a. Assessing grounding frequency using ship traffic and waterway complexity. J. Navigation 68 (1), 89–106.
Mazaheri, A., Montewka, J., Kujala, P., 2014. Modeling the risk of ship grounding – a literature review from a risk management perspective. WMU J. Maritime Affairs 13 (2), 269–297.
Mazaheri, A., Montewka, J., Nisula, J., Kujala, P., 2015b. Usability of accident and incident reports for evidence-based risk modeling – a case study on ship grounding reports. Saf. Sci. 76, 202–214.
Mccafferty, D.B., Baker, C.C., 2006. Trending the causes of marine incidents. In: 3rd Learning from Marine Incidents Conference. London, UK.
Merrick, J.R.W., van Dorp, J.R., 2006. Speaking the truth in maritime risk assessment. Risk Anal. 26 (1), 223–237.
Montewka, J., Ehlers, S., Goerlandt, F., Hinz, T., Tabri, K., Kujala, P., 2014a. A framework for risk assessment for maritime transportation systems—a case study for open sea collisions involving RoPax vessels. Reliability Eng. Syst. Saf. 124, 142–157.
Montewka, J., Goerlandt, F., Kujala, P., 2014b. On a systematic perspective on risk for formal safety assessment (FSA). Reliability Eng. Syst. Saf. 127, 77–85.
Nisula, J., 2014. Safety Factors in the 'tiedosta toimenpiteisiin' (TiTo) Project. Liikenteen Turvallisuusvirasto, Trafi, Helsinki.
O'hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T., 2006. Uncertain Judgements: Eliciting Experts' Probabilities. John Wiley & Sons, New York.
Oña, J.D., Mujalli, R.O., Calvo, F.J., 2011. Analysis of traffic accident injury severity on Spanish rural highways using Bayesian Networks. Accident Anal. Prevention 43, 402–411.
Özbaş, B., 2013. Safety risk analysis of maritime transportation: review of the literature transportation research record. J. Transportation Res. Board 2326, 32–38.
Pedersen, P.T., 1995. Collision and grounding mechanics. In: Proceedings of the Proceedings of WEMT '95', Copenhagen, Denmark.
Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 27 (8), 1226–1238.
Pitchforth, J., Mengersen, K., 2013. A proposed validation framework for expert elicited Bayesian networks. Expert Syst. Appl. 40 (1), 162–167.
Pitchforth, J., Wu, P., Mengersen, K., 2014. Applying a validation framework to a working airport terminal model. Expert Syst. Appl. 41, 4388–4400.
Pradhan, M., Henrion, M., Provan, G., Favero, B.D., Huang, K., 1996. The sensitivity of belief networks to imprecise probabilities: an experimental investigation. Artif. Intell. 85 (1–2), 363–397.
Psarros, G., Skjong, R., Eide, M.S., 2010. Under-reporting of maritime accidents. Accident Anal. Prevention 42 (2), 619–625.
Rambøll, 2006. Navigational Safety in the Sound Between Denmark and Sweden (øresund). Rambøll Danmark A/S, Denmark.
Rothbaum, A.R., 2000. Human Error and Marine Safety. National Safety Council Congress and Expo, Orlando.
Schröder-Hinrichs, J.U., Baldauf, M., Ghirxi, K.T., 2011. Accident investigation reporting deficiencies related to organizational factors in machinery space fires and explosions. Accident Anal. Prevention 43, 1187–1196.
SIAF, 2014. Marine investigation reports. Safety Investigation Authority of Finland.
Sormunen, O.V.-E., Goerlandt, F., Häkkinen, J., Posti, A., Hänninen, M., Montewka, J., Ståhlberg, K., Kujala, P., 2015a. Uncertainty in maritime risk analysis: extended case study on chemical tanker collisions. Proc. IMechE, Part M: J. Eng. Maritime Environ. 229 (3), 303–320.
Sormunen, O.V.-E., Hänninen, M., Häkkinen, J., Posti, A., 2015b. Tanker grounding frequency and spills in the Finnish Gulf of Finland. Sci. J. Maritime University Szczecin 43 (115), 108–114.
Steuer, R., Kurths, J., Daub, C.O., Weise, J., Selbig, J., 2002. The mutual information: detecting and evaluating dependencies between variables. Bioinformatics 18 (2), 231–240.
Theil, H., 1970. On the estimation of relationships involving qualitative variables. Am. J. Sociol. 76, 103–154.
Thomas, M., Skjong, R., 2009. Cost Benefit Analysis of inert gas systems for chemical and product tankers. In: Proceedings of the 28th International Conference on Ocean, Offshore and Arctic Engineering (OMAE), Honolulu.
Trochim, W., Donnely, J.P., 2008. The Research Methods Knowledge Base, third ed. Atomic Dog Publishing.
Valdezbanda, O.A., Goerlandt, F., Montewka, J., Kujala, P., 2015. A risk analysis of winter navigation in Finnish sea areas. Accident Anal. Prevention 79, 100–116.
van der Gaag, L.C., Renooij, S., Coupé, V.M.H., 2007. Sensitivity analysis of probabilistic networks. In: Lucas, P., Gámez, J.A., Salmerón, A. (Eds.), Advances in Probabilistic Graphical Models. Springer, pp. 103–124.
Wang, Y.F., Xie, M., Chin, K.-S., Fu, X.J., 2013. Accident analysis model based on Bayesian Network and evidential reasoning approach. J. Loss Prevent. Proc. 26, 10–21.
Wentworth, J.A., 2012. Verification, validation and evaluation of expert systems. In: A FHWA Handbook. U.S. Department of Transportation – Federal Highway Administration (FHWA).
Woods, D.D., 1988. Coping with complexity: the psychology of human behaviour in complex systems. In: Inc., T.F. (Ed.), Tasks, Errors, and Mental Models, pp. 128–148.
Zhang, W., Goerlandt, F., Montewka, J., Kujala, P., 2015. A method for detecting possible near-miss ship collision from AIS data. Ocean Eng. 107, 60–69.