Goerlandt, Floris; Montewka, Jakub

A framework for risk analysis of maritime transportation systems: A case study for oil spill from tankers in a shipship collision

Published in:
Safety Science

DOI:
10.1016/j.ssci.2015.02.009

Published: 01/01/2015

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

Safety Science 76 (2015) 42–66

Contents lists available at ScienceDirect

Safety Science

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

A framework for risk analysis of maritime transportation systems: A case study for oil spill from tankers in a ship–ship collision

Floris Goerlandt *, Jakub Montewka

Aalto University, School of Engineering, Department of Applied Mechanics, Marine Technology, Research Group on Maritime Risk and Safety, P.O. Box 15300, FI-00076 AALTO, Finland

ARTICLE INFO

Article history:
Received 13 August 2014
Received in revised form 12 January 2015
Accepted 12 February 2015
Available online 12 March 2015

Keywords:
Risk analysis framework
Ship–ship collision
Bayesian Network
Maritime safety
Risk analysis evaluation

ABSTRACT

This paper proposes a framework for risk analysis of maritime transportation systems, where risk analysis is understood as a tool for argumentative decision support. Uncertainty is given a more prominent role than in the current state of art in the maritime transportation application area, and various tools are presented for analyzing uncertainty. A two-stage risk description is applied. In the first stage, Bayesian Network (BN) modeling is applied for probabilistic risk quantification. The model functions as a communication and argumentation tool, serving as an aid to thinking in a qualitative evidence and assumption effect assessment. The evidence assessment is used together with a sensitivity analysis to select alternative hypotheses for the risk quantification, while the assumption effect assessment is used to convey an argumentation beyond the model. Based on this, a deliberative uncertainty judgment is made in the second risk analysis stage, which is supplemented with a global strength of evidence assessment. The framework is applied to a case study of oil spill from tanker collisions, aimed at response capacity planning and ecological risk assessment. The BN-model is a proactive and transferable tool for assessing the occurrence of various spill sizes in a sea area. While the case study uses evidence specific to the Gulf of Finland, the model and risk analysis approach can be applied to other areas. Based on evaluation criteria and tests for the risk model and risk analysis, it is found that the model is a plausible representation of tanker collision oil spill risk.

1. Introduction

In risk research, there is a recent focus on foundational issues. Calls have been made for devising risk analysis frameworks, focusing on issues such as how to understand and describe risk, and how to use risk analysis in decision making (Aven and Zio, 2014). Furthermore, there have been calls for devising methods for communicating uncertainty in risk analysis (Psaraftis, 2012).

In the maritime transportation application area, some theoretical frameworks exist, e.g. based on system simulation (Harrald et al., 1998), traffic conflict technique (Debnath and Chin, 2010) and Bayesian Networks (BNs) (Montewka et al., 2014a). Recent research has however shown that a wide range of definitions, perspectives and approaches to risk analysis co-exist, whereas typically little or no attention is given to risk-theoretic issues in applications. Furthermore, uncertainty typically is not considered (Goerlandt and Montewka, 2015).

In light of the above, this paper introduces a framework for risk analysis of maritime transportation systems, where uncertainty is given a more prominent role than in the current state of the art in the application area. Specific attention is given to the risk-theoretical basis (risk concept, perspective and use of risk analysis in decision making), and to the tools for analyzing uncertainties and biases beyond the model-based quantification. Bayesian Networks are applied as a modeling tool.

Subsequently, the framework is applied to a case study involving the oil spill risk from tankers in a ship–ship collision, aimed at providing insight in the possible occurrence of given spill sizes in this accident type in a given sea area. Such information is useful for response capacity and fleet organization planning (COWI, 2011; Jolma and Haapasaari, 2014; Lehikoinen et al., 2013) and for assessing the risk of biological impacts of oil spills (Lecklin et al., 2011).

While major oil spills from tankers are rare occurrences, the transportation of oil remains one of the main concerns for the various stakeholders in marine environmental protection (Dalton and Jin, 2010). This is due to their potentially major impact on marine ecosystems (Bi and Si, 2012), important socio-economic impacts on communities dependent on coastal resources (Garcia Negro et al., 2009; Miraglia, 2002) and high acute costs involved in clean-up operations (Montewka et al., 2013). Oil spills in harbor

* Corresponding author. Tel.: +358 9 470 23476; fax: +358 9 470 23493.
E-mail address: floris.goerlandt@aalto.fi (F. Goerlandt).

http://dx.doi.org/10.1016/j.ssci.2015.02.009
0925-7535/© 2015 The Authors. Published by Elsevier Ltd.

approaches or in narrow shipping waterways can also lead to a blockade, which can incur high costs to the world economy (Qu and Meng, 2012). Thus, adequate accident prevention measures (Hänninen et al., 2014; van Dorp and Merrick, 2011) and oil spill preparedness planning are important to enhance maritime safety and for marine environmental protection (IMO, 2010; Taylor et al., 2008).

Several methods and analyses have been proposed for assessing the oil spill risk from shipping activities in a sea area. Lee and Jung (2013) combine historic data with qualitative risk matrices for ranking likeliness and consequences. Quantitative methods for analyzing oil spill risk include event-trees and traffic flow theory or system simulation combined with ship collision damage modeling or accident statistics (Akhtar et al., 2012; COWI, 2011; Gucma and Przywarty, 2008; Li et al., 2012; Montewka et al., 2010b; van Dorp and Merrick, 2011). The work presented in this paper extends this literature by presenting a ship–ship collision oil spill risk analysis based on a Bayesian Network model.

This paper is organized as follows: In Section 2, a description of the applied understanding of the risk concept and the adopted two-stage risk perspective is given. A reflection is made on the intended use of the risk analysis. In Section 3, the methodological basis for the risk analysis framework is briefly outlined, focusing on the tools applied to describe risk. In Section 4, the case study to which the risk analysis framework is applied is introduced. The first risk analysis stage for the case study is presented in Section 5, and the second stage in Section 6. In Section 7, a discussion is given on the evaluation of the results, both concerning the risk model and the risk analysis as such. The utility of the tools for contextualizing the risk quantification is discussed in Section 8, and Section 9 concludes. For reasons of brevity, much of the data and models underlying the risk model and analysis is presented in Appendices.

2. Framework for risk analysis: risk-theoretic foundations

Risk analysis is an established tool for informing decisions. However, there are many different views on what risk is and how to define it (Aven, 2012; Hampel, 2006), how to measure/describe it (Aven, 2010a; Kaplan, 1997), and how to use risk analysis in decision making (Apostolakis, 2004; Aven, 2009). Therefore, this section provides a brief overview of the adopted conceptual understanding of risk, which systematic perspective is taken to describe risk and how to use the risk analysis results in decision making.

2.1. Risk concept: how risk is understood

Many definitions of the risk concept exist, involving constituents such as probability, uncertainty, possibility, events and/or consequences (Aven, 2012; Hampel, 2006). In the current application, risk is understood as referring to the possible but uncertain occurrence of a situation where something of human value is at stake. The terminological diversity of risk has philosophical roots, with opposing views on the nature of risk rooted in realism or constructivism (Shrader-Frechette, 1991). In the current understanding, risk is not considered a reality existing in itself, but a construct shared by a social group, informed by available evidence (Aven and Renn, 2009; Thompson and Dean, 1996). It is thus not a physical attribute of a system, existing by itself, but a concept attributed to a system in the mind of an assessor (Goerlandt and Montewka, 2015; Solberg and Njå, 2012).

2.2. Risk perspective: how risk is described

Understanding risk as above, a risk description is a reflection of a mind construct of analysts and experts (Aven and Guikema, 2011; Rosqvist, 2010; Watson, 1994), which may be more or less intersubjectively objective (Aven, 2010b). There is no reference to an underlying “true” risk, opposed to other risk description frameworks, such as the one presented by Kaplan (1997). In this section, the systematic manner to describe risk, i.e. the risk perspective, is outlined.

It is well-established that in the complex, distributed maritime transportation system, knowledge is not equally available about all parts of the system (Montewka et al., 2014b; Yan et al., 2014). Relying on poor evidence may lead to erroneous conclusions and misguided decisions, e.g. about risk acceptability or the choice between risk control alternatives. Because scientists have the responsibility to consider the consequences of error (Douglas, 2009), uncertainty has a central role in the current framework.

Moreover, in many analysis and modeling contexts, it is unavoidable to make simplifying assumptions which lead to conservative or optimistic biases in the analysis (Vareman and Persson, 2010). Such assumptions ultimately rely on value judgments (Diekmann and Peterson, 2013; Wandall, 2004). Because such value judgments may not be acceptable to all stakeholders (Hermansson, 2012), their effect is considered in the framework through considering biases.

Another issue is that the information presented to stakeholders and decision makers should be interpretable, i.e. it should be possible to explain what the presented numbers and descriptions mean (Aven, 2011a).

Based on the above, the current framework applies a two-stage risk description, as illustrated in Fig. 1. The general method and perspectives for analyzing risk is presented in the following sections. The intended use of the risk analysis in decision making is discussed in Section 2.3. The methodological aspects of the tools for analyzing risk are presented in Section 3.

2.2.1. Stage 1: Quantitative risk modeling with extended uncertainty/bias evaluation

In the first risk analysis stage, case-specific background knowledge is established. This concerns data, information, models and judgments and is conditional to a decision context, which sets limits to the available resources (time, money, expertise) and may act as a guide to selecting conservative or optimistic modeling choices (Vareman and Persson, 2010).

Using the evidence, a risk model is constructed to quantify risk, using the established background knowledge. This is supplemented with an extended qualitative assessment of uncertainties and biases. Together with a sensitivity analysis, model variables for which to prioritize alternative hypotheses are identified and their effect quantified using the risk model. The effect of assumptions on the quantitative risk results is finally qualitatively assessed.

The perspective applied in the first stage of the risk analysis is based on a combination of the precautionary perspective (Rosqvist and Tuominen, 2004) and the uncertainty-based perspective (Aven, 2010a; Aven and Zio, 2011; Flage and Aven, 2009). In the precautionary perspective, frequentist probability P_f and subjective probability P_s are used to describe events and consequences. This is supplemented by a qualitative assessment of model biases B, i.e. whether conservative or optimistic modeling choices are made. In the uncertainty-based perspective, the occurrence of events and consequences is quantified using subjective probabilities P_s. Structural uncertainty is quantified using alternative hypotheses in the model construction U_{AH} and/or through a qualitative assessment of uncertainties U_{QL}.

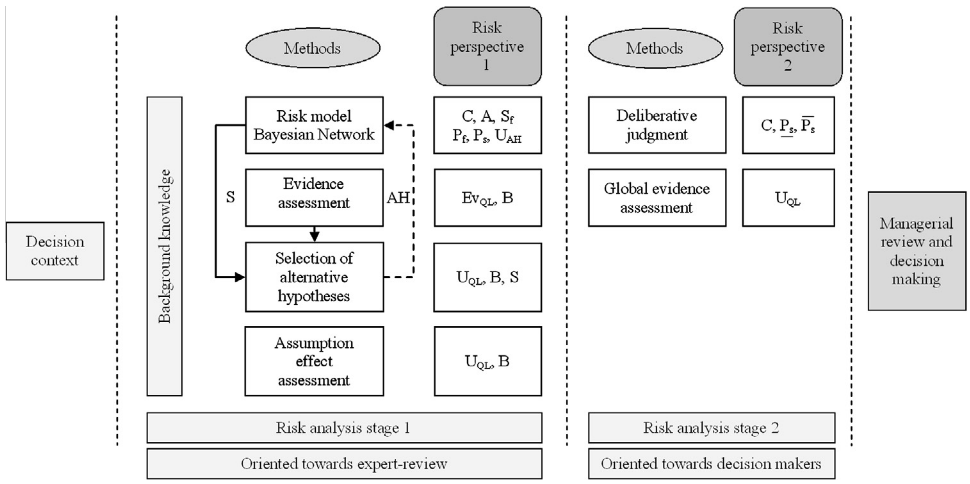

Fig. 1. 2-Stage framework for risk analysis: methods and perspectives.

The elements of the adopted risk perspective can be summarized as follows (where "∼" signifies "is described by")

$$R_1 \sim (C, E, S_f, P_f, P_s, U_{AH}, S, Ev_{QL}, B|BK) \quad (1)$$

The focus is on consequences C, the occurrence of which is conditional to events E. Instead of focusing on the events per se, their occurrence can also be described through a set of situational factors S_f. As measurement tools, frequentist and subjective probabilities P_f and P_s are used. Uncertainties and biases underlying the risk model construction are qualitatively assessed using an evidence assessment scheme (Ev_{QL} and B). Together with a sensitivity analysis S on the probabilistic risk model, this evidence assessment is used to select alternative hypotheses in the risk model, the effect of which are quantified using the risk model (U_{AH}). The risk analysis is conditional to a specific background knowledge BK, which consists of data, information, models, judgments and assumptions.

2.2.2. Stage 2: Deliberation, judgment-based quantification and qualification based on results obtained from Stage 1

In the second risk analysis stage, the information obtained from the first stage is used to quantify the uncertainty about the occurrence of consequences C using interval probabilities $\underline{P_s}$ and $\overline{P_s}$ and a qualitative assessment of the underlying evidence U_{QL}:

$$R \sim \left(C, \underline{P_s}, \overline{P_s}, U_{QL}|R_1\right) \quad (2)$$

The probability interval $[\underline{P_S}, \overline{P_s}]$ is a judgment of a (group of) expert(s), expressing imprecision regarding the uncertainty about the occurrence of consequences C. These can be interpreted by reference to an uncertainty standard, i.e. as a lower and upper bound of the degree of belief of drawing a particular ball from an urn (Aven, 2011a; Lindley, 2006). These judgments are based on the evidence as obtained from the first risk analysis stage, in particular the probabilistic results including the quantitative uncertainty bounds using the alternative hypotheses, the qualitative evidence assessment and the assumption effect assessment.

2.3. Use of risk analysis in decision making

An important issue is how the risk analysis is meant to be applied in decision making. Some frameworks use risk analysis as a tool for calculating probabilities, which are compared with a risk acceptability criterion for making the decision in a quasi-automatic manner, see e.g. de Rocquigny et al. (2008). Other frameworks codify the value judgments over the outcomes through a utility function, and may apply a form of mathematical optimization to select the proposed decision, see e.g. Kaplan (1997). Examples of such views on risk analysis in the context of maritime oil spills are found in e.g. Klanac and Varsta (2011) and Lehikoinen et al. (2013).

The framework presented here does not focus on the probabilities per se, i.e. these are not used to 'prove' that the risk is acceptable or that a certain risk management action should be performed. The aim is rather to concisely communicate evidence from analysts and experts to decision makers, which is used further in a broad decision making process (Aven, 2009), where other aspects relevant to the decision, e.g. the availability of resources and strategic or socio-economic concerns are considered. Thus, risk analysis is understood as risk-informed, not risk-based (Apostolakis, 2004).

This follows from the envisaged functions of the risk model constructed in the first risk analysis stage, namely to (i) convey an argumentation based on available evidence, (ii) provide a basis for communication between stakeholders, and (iii) serve as an aid to thinking. Such functions are acceptable for non-predictive models (Hodges, 1991). Thus, the risk model does not in itself lead to a risk characterization, but is essentially connected with the qualitative evidence and assumption effect assessments. The purpose of these qualitative assessments is to moderate the argument made by the risk quantification using the model, and to provide transparency about the risk analysis and its underlying evidence, which are key aspects of risk-informed decision making (Aven, 2011b; Watson, 1994).

The first risk analysis stage leads to a quite elaborate characterization, which aims to provide full transparency about the risk analysis. One possible challenge of this extensive assessment is that decision makers may in practical settings not have time to go through all the material. Hence, the primary intended users of the first analysis phase are a panel of expert-reviewers, such as the FSA[1] Expert Group in IMO[2] decision making (Psaraftis, 2012). The second risk analysis stage provides a simplified and concise insight in the risk and the strength of evidence, as a kind of summary of the findings of the first stage. Thus, the intended users of this analysis phase are the actual decision makers, who likely do not have the expertise nor the time to review the complete risk analysis.

3. Framework for risk analysis: methodological aspects

In this section, the methodological aspects of the framework of Fig. 1 are presented, i.e. the tools for measuring risk. The following

[1] FSA: Formal Safety Assessment.
[2] IMO: International Maritime Organization.

items are addressed concerning the first risk analysis phase: (i) Bayesian Networks as a tool for propagating uncertainty, including its functionality for parameter sensitivity analysis, (ii) the alternative hypotheses approach for accounting for epistemic uncertainty, (iii) the method for qualitatively assessing the evidence base, (iv) the procedure for selecting alternative hypotheses and (v) the method for assessing the effect of assumptions on the model output. For the second risk analysis phase, the method for global uncertainty evaluation is shown.

3.1. Bayesian Networks

Bayesian Networks are selected as a risk modeling tool as these have a number of favorable characteristics. BNs can contextualize of the occurrence of specific consequences through situational factors, which represent observable aspects of the studied system. They furthermore allow integration of different types of evidence through various types of probabilities and provide a means for performing sensitivity analysis. It is also rather straightforward to incorporate alternative hypotheses in the model. Bayesian Networks are relatively widely used tools for risk modeling (Aven, 2008; Fenton and Neil, 2012).

3.1.1. Mathematical basis

BNs represent a class of probabilistic graphical models, defined as a pair $\Delta = \{G(\boldsymbol{V}, \boldsymbol{A}), \boldsymbol{P}\}$ (Koller and Friedman, 2009), where $G(\boldsymbol{V}, \boldsymbol{A})$ is the graphical component and $\boldsymbol{P}$ the probabilistic component of the model. $G(\boldsymbol{V},\boldsymbol{A})$ is in the form of a directed acyclic graph (DAG), where the nodes represent the variables $\boldsymbol{V} = \{V_1, \ldots, V_n\}$ and the arcs ($\boldsymbol{A}$) represent the conditional (in)dependence relationships between these. $\boldsymbol{P}$ consists of a set of conditional probability tables (CPTs) $P(V_i|\boldsymbol{Pa}(V_i))$ for each variable V_i, $i = 1, \ldots, n$ in the network. $\boldsymbol{Pa}(V_i)$ signifies the set of parents of V_i in G: $\boldsymbol{Pa}(V_i) = \{Y \in \boldsymbol{V} | (Y, V_i) \in \boldsymbol{A}\}$. Thus: $\boldsymbol{P} = \{P(V_i|\boldsymbol{Pa}(V_i)), i = 1, \ldots, n\}$. A BN encodes a factorization of the joint probability distribution (JDP) over all variables in $\boldsymbol{V}$:

$$P(V) = \prod_{i=1}^{n} P(V_i|\boldsymbol{Pa}(V_i)) \tag{3}$$

3.1.2. Interpretation of BNs in context of risk analysis

In the risk analysis, the variables $\boldsymbol{V}$ may represent the events $\boldsymbol{E}$, situational factors $\boldsymbol{S_f}$, the consequences $\boldsymbol{C}$, the nodes for weighing the alternative hypotheses $\boldsymbol{AH}$ and the risk metrics $\boldsymbol{AR}$. The arcs $\boldsymbol{A}$ represent the relations between these. The CPTs for the situational factors $\boldsymbol{S_f}$ are derived from the evidence base. These consist of frequentist probabilities P_f if a population of similar conditions can be meaningfully defined and if the knowledge base contains data or models from which relative frequencies can be calculated, and of subjective probabilities P_s if these are judgments by an assessor (Aven and Reniers, 2013). Thus, P_f can be understood to describe aleatory uncertainty, representing the inherent variation in a system. P_s can be understood as a degree of belief, a measure of epistemic uncertainty, which results from lack of knowledge (Faber, 2005). The CPTs for $\boldsymbol{AH}$ are based on judgments of an assessor, and are best understood as a weight of credibility of a model to appropriately reflect the underlying mechanisms the model attempts to cover (Aven, 2010c).

3.1.3. Sensitivity analysis

The purpose of a sensitivity analysis is to investigate the effect of changes in the assigned probabilities of the network variables on the probabilities of a specific outcome variable. In a one-way sensitivity analysis, every conditional and prior probability in the network is varied in turn, keeping the others unchanged. A sensitivity-value approach presented by Coupé and van der Gaag (2002) is applied. A sensitivity function is defined which describes a specific risk metric RM as a function of the parameter $z = p(RM = RM_i|\boldsymbol{\pi})$, where RM_i is one state of the risk metric, and $\boldsymbol{\pi}$ is a combination of states for the parent nodes of RM. For a network with no observations on any of the network variables, a linear sensitivity function is found:

$$RM(z) = u_1 z + u_2 \tag{4}$$

where the constants u_1 and u_2 are identified based on the model. The first derivative of the sensitivity function at the base value describes the effect of minor changes in the original parameter value on the value of the output, leading to a numerical sensitivity value:

$$RM'(z) = u_1 \tag{5}$$

The sensitivity of a BN-variable V on the risk metric RM is considered by $\max|u_1|$.

3.2. Alternative hypothesis approach

If the evidence base for a situational factor is poor and if alternative plausible alternatives are available, one theoretical methodology to consider evidential uncertainty is to apply alternative hypotheses (Zio and Apostolakis, 1996). The rationale can be summarized as follows. Consider EB_i one alternative from the evidence base. Conditional to EB_i, probabilities for a situational factor S_f are derived: $P_i(S_f|EB_i)$, for $i = 1, \ldots, n$. These probabilities are weighed using subjective probabilities p_i, with $\sum_i^n p_i = 1$. Weighed probabilities are obtained as follows:

$$P = \sum_{i}^{n} P_i(S_f|EB_i) p_i \tag{6}$$

Thus, the effect of alternative hypotheses on the probabilities over the consequence space can be quantified, providing insight in the stability of the results in light of uncertain evidence.

3.3. Qualitative multi-criteria evidence assessment

The qualitative assessment of the evidence base, which communicates which elements of the model are based on strong or poor evidence, is performed using an adapted version of a method proposed by Kloprogge et al. (2011). Various aspects of each data source or model are rated by an assessor using a qualitative scale, as shown in Table 1. The ratings range from 1 to 5, with intermediate values for conditions between the extremes. These are subjective judgments by an assessor. They are intended to contextualize the quantitative argument put forward by the risk model, increasing the transparency about its evidence base. The assessment also has a role in prioritizing the selection of alternative hypotheses.

Three types of evidential characteristics are considered. First, the knowledge dimension addresses how strong the evidence base is, and concerns the quality, completeness and amount of available data and the empirical adequacy and theoretical viability of models. Second, the influence of value-laden choices is considered through evaluating the direction of bias, assessing how conservative or optimistic the model output is compared with its unknown true value. Third, the contextual dimension considers the choice space and the influence of contextual limitations on the modeling choices made by the analyst. These characteristics are interpreted as in Table 1.

Table 1
Assessment scheme for evidence qualities.

Evidence base	Score = 1	Score = 5
Data		
Quality	High number of errors Low accuracy of recording Low reliability of data source	Low number of errors High accuracy of recording High reliability of data source
Amount of data	Little available data	Much relevant data available
Completeness	High number of missing data fields High level of underreporting	Low number of missing data fields Low level of underreporting
Models		
Empirical validation	No experimental confirmation available	Many different experimental tests performed
	Existing experimental tests show large discrepancy with model output	Existing experimental tests agree well with model output
Theoretical viability	Model expected to lead to poor predictions	Model is expected to provide good predictions
Bias	The model leads to optimistic predictions compared with the real values	The model leads to conservative predictions compared with the real values
All types		
Contextual limitations	Totally different BK would have been used, were more time/money/ resources available	The same BK would have been used, irrespective of available time/money/ resources available
Choice space	Few alternatives available in BK	Many alternatives available in BK

3.4. Evidence importance ranking and selection of alternative hypotheses

As outlined in Section 3.2, alternative hypotheses can be included in the risk model. Even though several authors have suggested this method (Aven and Zio, 2011; Zio and Apostolakis, 1996), no guidance is provided as to which model elements to prioritize for considering alternative hypotheses.

In the current framework, this is done by tabulating the strength of evidence and the direction of bias along with the results of the sensitivity analysis performed on the Bayesian Network, as illustrated in Fig. 2. The interpretation of the ratings is shown in the figure as well, based on ideas presented in Flage and Aven (2009) and Rosqvist and Tuominen (2004).

Situational factors with high sensitivity and low evidential support and/or strongly biased evidence (shown in red[3]) are considered more important from a decision maker's point of view. This is because poor evidence can lead to poorly justified probability assignments, while the analysis outcome is sensitive to changes in the parameterization. If biased variables significantly affect the outcome, this is important to consider as the aggregate effect of various value judgments may not be in line with stakeholder values. Thus, model variables with high importance scores in Fig. 2 are prioritized for selecting alternative hypotheses.

3.5. Assessment of effect of assumptions

In risk analyses, assumptions often constitute an important part of the background knowledge, and it is typically not possible to account for all assumptions by considering alternative hypotheses. In the current risk perspective, assumptions have a prominent role. Considering the role of the risk model as a tool for argumentation, a means of communication and a platform for thinking, assumptions provide focal points in the deliberation about the results of the risk model as obtained from the first stage in the analysis. These assumptions may relate to the relations between the model variables, the parameterization of the probability tables underlying the variables or address factors not included in the model.

The information obtained from an assumption effect assessment is used in the second stage of the risk analysis, through a deliberative process where it is assessed whether the assumptions are plausible, and if not, how the risk model results would be affected.

To evaluate the influence of assumptions, an assessment method is developed, adapting ideas presented by Aven (2013). The assumption is evaluated by considering (i) the magnitude of the deviation due to the incorrectness of the assumption, (ii) the direction of this deviation, (iii) consequence range where the deviation has effect and (iv) the strength of justification for making the assumption assessment. A crude, qualitative assessment is applied for these characteristics, see Table 2.

3.6. Global assessment of evidence

In the second risk analysis stage, a global evidence assessment accompanies the risk quantification. This is performed using a crude scheme proposed by Flage and Aven (2009). A direct grading of the importance of the uncertainty is performed through a judgment of an assessor of four criteria. A justification for the assessment of each criterion can be provided. The strength of the evidence underlying the risk quantification can be visualized using a color code, with strong, medium and poor quality of evidence represented by green, yellow and red. The assessment scheme uses a classification as in the "strength of justification" of Table 2.

4. Case study: introduction

The aim of the risk analysis is to provide insight in the possibility of occurrence of oil spills from collisions with oil tankers. The focus is on how likely spills of certain sizes are expected to occur. This information is useful for response capacity planning and response fleet organization as well as for assessing the biological impacts of oil spills (COWI, 2011; Helle et al., 2011; Lehikoinen et al., 2013).

In the application presented here, the considered area is the part of the Gulf of Finland covered by the GOFREP[4] system and the Russian national VTS[5] sector, see Fig. 3. This area contains most of the traffic in the area, as ships are required to follow the sea lanes of the Traffic Separation Scheme (TSS). As the current response capacity (30,000 tonnes) is based on a plausible worst case accident (Jolma and Haapasaari, 2014), a precautionary decision maker is assumed. Thus, where required in the analysis, conservative assumptions are preferred over optimistic ones.

5. Case study: risk analysis stage 1

In this section, the first stage of the risk analysis framework presented in Sections 2 and 3 is shown for the case study.

5.1. Bayesian Network model

The structure of the BN-model is defined in Fig. 4. The focus of the analysis is the collision consequence, i.e. the amount of oil

[3] For interpretation of color in Fig. 2 and 4, the reader is referred to the web version of this article.

[4] GOFREP: Gulf of Finland reporting system.

[5] VTS: Vessel Traffic Service.

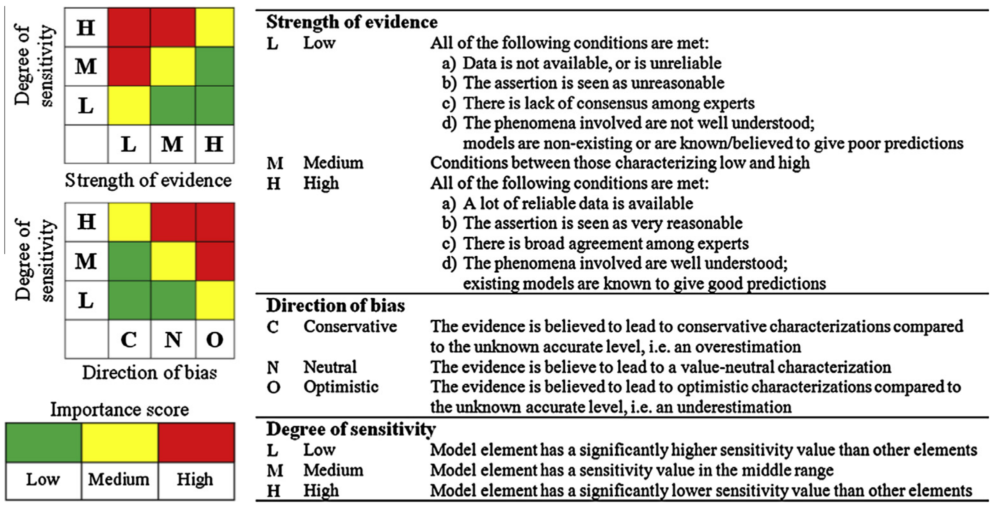

Fig. 2. Evidence assessment scheme for selecting alternative hypotheses.

Table 2
Assessment scheme for assumption effect.

Magnitude of deviation		
L	Low	Maximum plausible changes in base values result in outcome changes less than an order of magnitude
M	Medium	Maximum plausible changes in base values result in outcome changes of about an order of magnitude
H	High	Maximum plausible changes in base values result in outcome changes of two orders of magnitude or more
Direction of deviation		
+	Increase	The deviations result in an outcome which is higher than the risk model suggests
±	Increase/ decrease	The deviations result in an outcome which is higher or lower than the risk model suggests
–	Decrease	The deviations result in an outcome which is lower than the risk model suggests
Consequence range where deviation has effect		
L	Low	Deviations have effect in lower third of the consequence range considered in the quantification
M	Medium	Deviations have effect around the middle of the consequence range considered in the quantification
H	High	Deviations have effect in the upper third of the consequence range considered in the quantification
B	Beyond maximum	Deviations have effect beyond the maximum quantitatively considered consequence range
Strength of justification		
L	Low	All of the following conditions are met: (a) Data is not available, or is unreliable (b) The assertion is seen as unreasonable (c) There is lack of consensus among experts (d) The phenomena involved are not well understood; models are non-existing or are known/believed to give poor predictions
M	Medium	Conditions between those characterizing low and high strength of justification
H	High	All of the following conditions are met: (a) A lot of reliable data is available (b) The assertion is seen as very reasonable (c) There is broad agreement among experts (d) The phenomena involved are well understood; existing models are known to give good predictions

spilled. The occurrence of a spill of a certain size is conditional to two situations: encounter and impact between vessels.

The encounter situation addresses the conditions under which tankers encounter other vessels in the area. This is described through a set of situational factors, including the location of encounter in the sea area, direction of travel, encounter type, and characteristics of the encountering vessel (ship type, length, mass, encounter speed, bow angle and loading condition) and tanker (length, width, mass, deadweight, encounter speed and loading condition). These situational factors are included as model variables $\boldsymbol{S_f^E}$. In Fig. 4, the green nodes relate to variables of the encountering vessels, the light blue nodes of the tankers and the pink nodes of the encounter conditions between both vessels.

The impact situation addresses the conditions under which tankers may collide with other vessels. As vessels typically perform evasive action prior to collision (Cahill, 2002; Wang et al., 2013), impact conditions differ from conditions at encounter. Impact situations are described using a set of situational factors, conditional and/or additional to the encounter situational factors. These include both vessels' impact speeds, the impact angle, the location of impact along the hull and the probability that the tanker is the struck vessel. The occurrence of a collision where a tanker is the struck vessel is also considered in the impact situation, conditional to the occurrence of a collision and the event that the tanker is the struck vessel. These situational factors are included as model variables $\boldsymbol{S_f^I}$. In Fig. 4, these are the yellow nodes.

Conditional to the impact conditions, a plastic deformation occurs in the contact area between the two vessels, resulting in damage to the struck vessel's hull. Depending on location and extent of this damage, cargo or bunker oil tanks may be breached, resulting in a spill. The occurrence of a hull breach and subsequent oil spill is the consequence *C*. In Fig. 4, alternative models for the damage extent and spill are grouped in the orange node.

The risk metric RM is the discrete probability distribution describing the occurrence of various spill sizes, indicated in dark gray in Fig. 4. The model also contains alternative hypotheses, denoted as model variables ***AH***, indicated in light gray in Fig. 4.

5.2. Outline of the evidence base

In this section, a brief outline is given of the evidence base used to calculate or assess probabilities for the probability tables underlying the Bayesian Network. This outline is made on a generic level, focusing on the data and models required for the parameterization, see Table 3. This is done for reasons of brevity and to make the presented BN-model more accessible for possible users concerned with other sea areas.[6] Reference is made to Appendix A, where the

[6] One benefit of BNs is that the parameterization of the probability tables can in principle be done using different types of evidence. In particular, where data or models are not available to derive probabilities for a specific sea area, the probability tables can be populated using knowledge-based probabilities P_s.

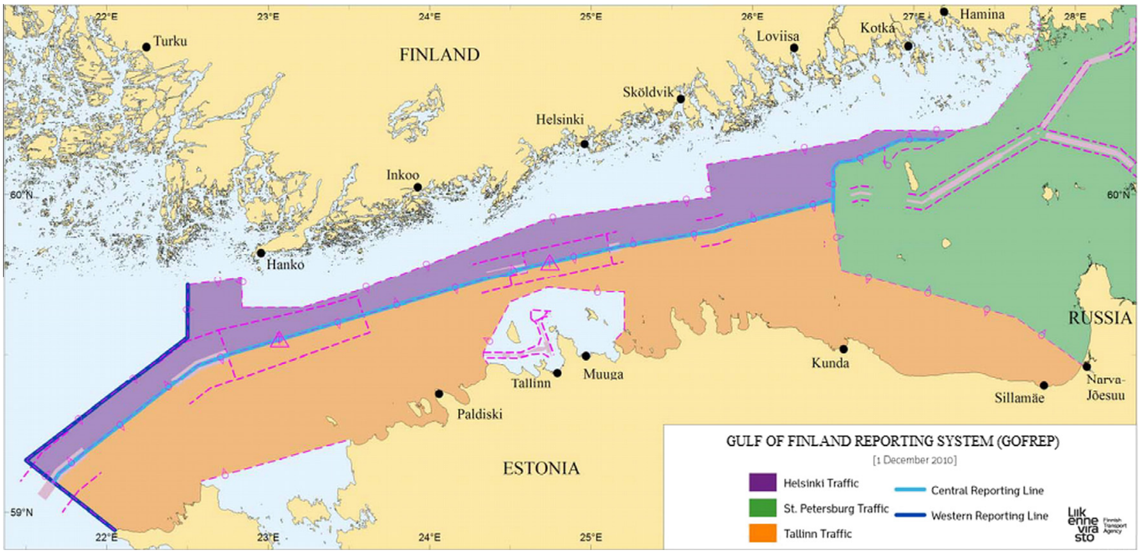

Fig. 3. Definition of the considered area, adapted from FTA (FTA, 2010).

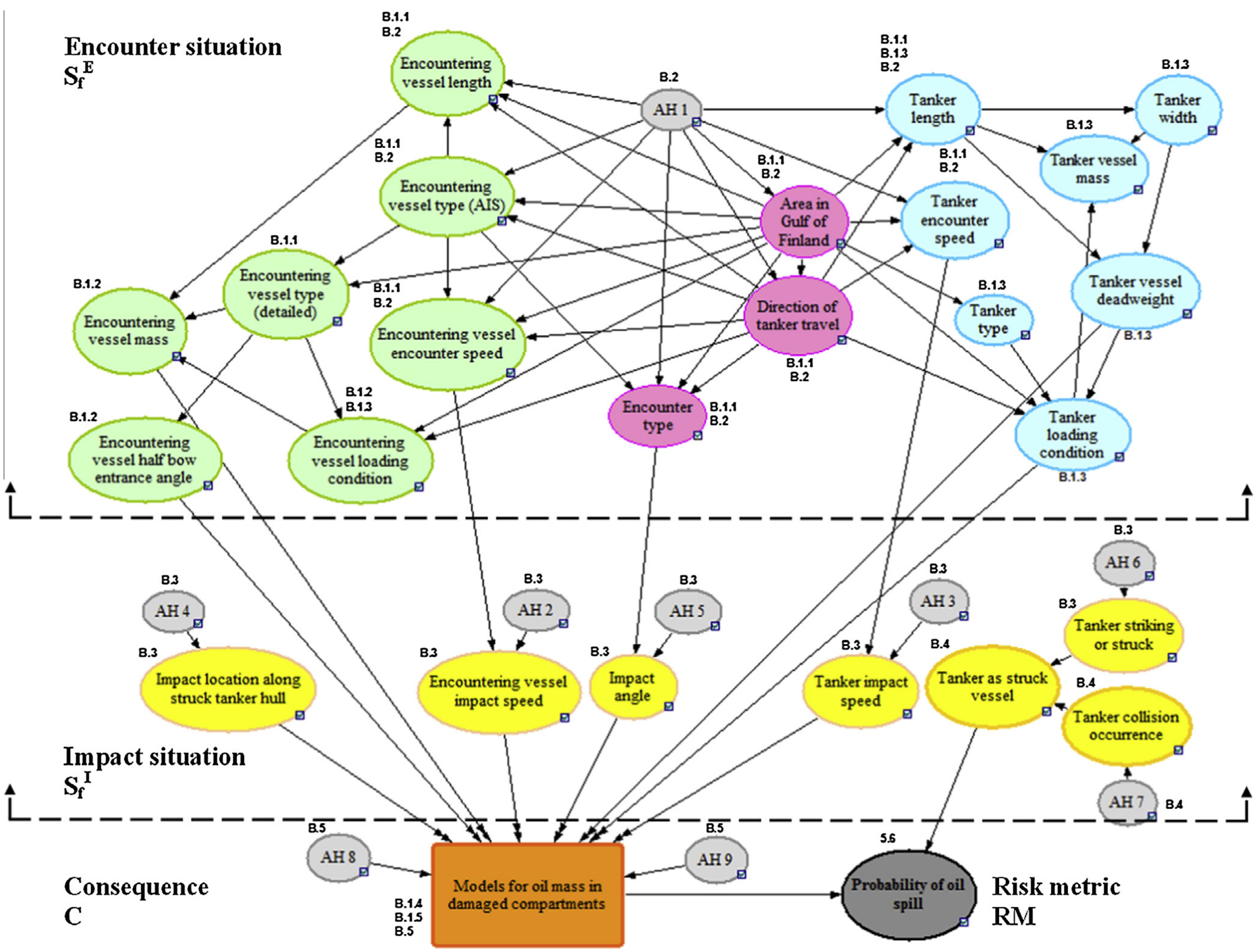

Fig. 4. Structure of the oil spill risk BN-model, the labels next to the nodes refer to the sections where the supporting evidence is described.

BN-variables and their discretization is listed. Appendix B summarizes the evidence applied in this case study, providing deeper insight in the specific meaning of the required evidence and how these are combined to derive probabilities.

5.3. Qualitative multi-criteria evidence assessment

The evidence is qualitatively assessed using the rating scheme presented in Section 3.3. Results are shown in Tables 4 and 5. As

Table 3
Outline of evidence underlying the BN model.

Factor group Variable		Required evidence (data, models)	Ref.
Encounter situation – tankers (S_f^E)			
V_1	Tanker length	AIS data, encounter detection method, *alternative hypothesis*	B.1.1, B.1.3, B.2
V_2	Tanker width	Ship characteristics database	B.1.3
V_3	Tanker mass	Ship characteristics database, tanker tank configuration models	B.1.3, B.1.4
V_4	Tanker encounter speed	AIS data, encounter detection method, *alternative hypothesis*	B.1.1, B.2
V_5	Tanker type	AIS data, encounter detection method, cargo flow analysis	B.1.3
V_6	Tanker deadweight	Ship dimensions database	B.1.3
V_7	Tanker loading condition	Cargo flow analysis	B.1.3
Encounter situation – encountering vessel (S_f^E)			
V_8	Encountering vessel length	AIS data, encounter detection method, *alternative hypothesis*	B.1.1, B.2
V_9	Encountering vessel mass	Ship characteristics database, ship dimension regression models	B.1.2, B.1.3
V_{10}	Encountering vessel type (AIS)	AIS data, encounter detection method, *alternative hypothesis*	B.1.1, B.2
V_{11}	Encountering vessel type (detailed)	Cargo flow analysis	B.1.1
V_{12}	Encountering vessel encounter speed	AIS data, encounter detection method, *alternative hypothesis*	B.1.1, B.2
V_{13}	Encountering vessel loading condition	Cargo flow analysis	B.1.2, B.1.3
V_{14}	Encountering vessel bow entrance angle	Ship characteristics database, ship dimension regression models	B.1.2
Encounter situation – context (S_f^E)			
V_{15}	Area in Gulf of Finland	AIS data, encounter detection method, *alternative hypothesis*	B.1.1, B.2
V_{16}	Direction of tanker travel	AIS data, encounter detection method, *alternative hypothesis*	B.1.1, B.2
V_{17}	Encounter type	AIS data, encounter detection and classification method, *alternative hypothesis*	B.1.1, B.2
Impact situation (S_f^I)			
V_{18}	Encountering vessel impact speed	Accident data, expert judgment, *alternative hypothesis*	B.3
V_{19}	Impact angle	Accident data, expert judgment, *alternative hypothesis*	B.3
V_{20}	Tanker impact speed	Accident data, expert judgment, *alternative hypothesis*	B.3
V_{21}	Impact location along struck tanker hull	Accident data, expert judgment, *alternative hypothesis*	B.3
V_{22}	Tanker striking or struck	Accident data, expert judgment, *alternative hypothesis*	B.3
V_{23}	Tanker collision occurrence	Accident data, accident underreporting studies AIS data, *alternative hypothesis*	B.4
V_{24}	Tanker as struck vessel	–	B.4
Consequences – oil spills from damaged compartment (C)			
V_{25}–V_{29}	Oil spill mass	Collision damage extent model, tank layout model, oil outflow model, bunker tank layout model, bunker tank size data, *alternative hypotheses*	B.5, B.1.3, B.1.4, B.1.5

an illustration of the feasibility of justifying the ratings, some examples of elements in the evidence are considered below. The data and models referred to are more elaborately described in Appendix B.

The data assessment in Table 4 indicates that the data used for describing the encounter situation is generally good. For instance, the applied AIS data (D1) is very extensive: over 5 months of high-resolution data is used, covering the entire study area. AIS data quality has been mediocre in earlier years (Graveson, 2004), but has improved significantly (Felski and Jaskolski, 2013). For the bow angle of encountering vessels (D3), only very limited data is available. The quality is good, but it is based on a limited review of ship drawings. It is uncertain in how far this data is representative to the vessels operating in the studied area. It would in principle be possible to repeat such a review of drawings for vessels in the area, i.e. there are alternative ways to gather evidence. However, due to contextual limitations, this is not considered a high priority.

The data concerning the layout and size of the cargo tanks (D5) is considered extensive, of good quality and accurately reflecting the conditions in the studied area. However, there is only very limited data for the bunker tanks available (D6), with low quality. As this data is gathered for ships operating in the waters of the United States, it is not known in how far the bunker data is representative for vessels operating in the Gulf of Finland. However, the exact bunker tank sizes are less important than the cargo tank sizes, as the former concern the lower consequence ranges and because of the crudeness of the applied discretization of the consequence variable, see Table 4. Thus, improving the bunker tank data is possible e.g. by reviewing ship drawings, but this is not considered a priority to improve the analysis. This is generally the case for the data: better data may exist, but priority for improving the analysis lays elsewhere.

The assessment of the applied models in the analysis shows that most models have not, or not extensively, been empirically validated, see Table 5. The model for the tank sizes (M5) has been compared with a set of tanker designs typical for the Baltic Sea, showing good agreement (Smailys and Česnauskis, 2006). The regression model for mass of encountering vessels (M2) is also validated through application of statistical tests (Brown, 2002), with good agreement. For tankers, the model for tank sizes is applied to derive a mass, and is validated as mentioned above. The damage extent model (M6) is compared with a limited number of damage cases, indicating conservative damage estimates (Chen, 2000). This is mainly due to the application of the collision damage model to midship sections, whereas it is known that the hull structural capacity near bulkheads is greater than near the midpoint between bulkheads (Klanac et al., 2010).

Most models are considered theoretically plausible for the aims of the risk analysis. The encounter detection method (M1) requires the specification of an inspection domain. Many choices are possible, but the circular domain is considered reasonably plausible to derive the characteristics of vessels which tankers encounter in the area. The encounter type classification model (M3) is considered justified as it follows the rationale of the COLREGs.[7] The statistical damage extent model (M6) is based on damage cases calculated using a three degree of freedom time-domain simulation model (Brown and Chen, 2002; van de Wiel and van Dorp, 2011). It is less accurate than state-of-the art models for ship design purposes (Ehlers and Tabri, 2012), but because it explicitly accounts for the coupling of outer and inner dynamics, it outperforms decoupled approaches especially for non-perpendicular impacts (Tabri, 2010).

[7] COLREGs: Collision regulations, i.e. Convention of the International Regulations for Preventing Collisions at Sea.

Table 4
Assessment of qualities of data underlying the analysis, using the rating scheme of Table 1.

Characteristic	D1	D2	D3	D4	D5	D6	D7
Quality	4	5	4	4	5	1	5
Amount of data	5	4	1	5	5	1	1
Completeness	4	4	2	4	5	1	5
Contextual limitations	5	5	3	5	5	4	5
Choice space	2	2	3	1	2	1	1

Label	Data	Section
D1	Data Automatic Information System	B.1.1
D2	Data from traffic flow analysis	B.1.1, B.1.3
D3	Bow angle of encountering vessels	B.1.2
D4	Loading condition oil tanker	B.1.3
D5	Tanker tank layout	B.1.3
D6	Bunker tank size and layout	B.1.5
D7	Collision probability	B.4

As many impact models lead to non-perpendicular impacts, see Table B7, this is a desirable model feature. The statistical meta-model for the damage extent is reported with a good statistical fit for the variables x_i, based on an assessment of probability plots, which show an overprediction of the damage sizes. The quality of the regression models of Eq. (B7) and Eq. (B8) is good with reported R^2-values of 70.6% for y_L and 73.6% for y_T and also the model for the damage direction θ is reported with good statistical fit with the damage cases reported by NRC (2001). However, the damage extent model is based on a limited number of ship designs, whereas it is known that the specific structural design of the struck ship's hull has an influence on the collision damage (Hogström, 2012; Klanac et al., 2010). Moreover, the damage extent model does not account for yaw and sway velocities at the moment of impact, even though it is known that these affect the collision energy (Ståhlberg, 2010) and damage extent (Wiśniewski and Kołakowski, 2003). The oil spill model (M7) is very crude. The assumption that all oil is spilled is plausible if the tanker is struck near and below the waterline. However, for double hull tankers, the ballast tank will retain some oil and depending on the opening size and damage height, the duration of the oil outflow can take several hours or even days. Hence, not necessarily all oil from a tank is spilled (Tavakoli et al., 2010). This supports the assessment that M7 is conservative, i.e. that the actual oil outflow will be lower.

Both M6 and M7 are conservative models, which is acceptable for precautionary decision-making. However, their conservativeness precludes a more detailed insight in the oil spill sizes. Considering the assessment of the contextual limitations, it is thus found most beneficial to use more detailed damage extent models (M6) and oil outflow models (M7) to improve the analysis.

Table 5
Assessment of qualities of models underlying the analysis, using the rating scheme of Table 1.

	M1	M2	M3	M4	M5	M6	M7
Empirical validation	1	3	1	1	3	2	1
Theoretical viability	3	3	5	2	4	3	2
Bias	3	5	3	3	3	5	5
Contextual limitations	5	3	5	4	5	1	1
Choice space	5	2	3	5	1	3	3

Label	Model	Section
M1	Encounter detection model	B.2
M2	Encounter vessel mass model	B.1.2
M3	Encounter type classification model	B.2
M4	Models for collision impact variables	B.3
M5	Tank volume model	B.1.4
M6	Collision damage model	B.5.1
M7	Oil spill model	B.5.2

5.4. Evidence importance ranking and selection of alternative hypotheses

The parameter sensitivity analysis is performed on the BN with the alternative hypotheses listed in Table A1 (Appendix A) set at their most likely value, or at the first alternative for equiprobable cases. Application of the procedure in Section 3.1.3 shows that the model is most sensitive to changes in variables listed in Table 6. It is seen that the model output is most sensitive to the probabilities assigned to the variables describing whether a collision involving a tanker occurs, and whether it is the struck vessel. Furthermore, the model is sensitive to the oil spill mass in the damaged tanks and to variables describing the struck tanker, in particular the deadweight, loading condition and impact location along the struck hull and impact angle. The output is somewhat sensitive to parent variables to these situational factors: the direction of tanker travel and the area in the Gulf of Finland affect the loading condition and tanker length, whereas the tanker length and width affect the tanker deadweight.

The strength and bias of the evidential support are further mapped in relation to the parameter sensitivity, providing insight in the importance of the weaknesses of the evidence base and the priorities of evaluating the effect of alternative hypotheses on the quantitative results. This is performed in Fig. 5, for a precautionary decision maker (i.e. preferring conservative modeling choices over optimistic ones). For each variable, it is considered which data and/or models underlie the probability assignments, as indicated in the graphical BN-representation of Fig. 4. This in turn is used to inspect the evidence assessment scores in Tables 4 and 5, from which a rating of low, medium or high strength of evidential support is obtained as in Fig. 2. Similarly, the ratings of the evidence base lead to a rating in terms of conservative, neutral or optimistic bias underlying the variable parameterization.

The rather rough assessment indicates that the overall risk analysis is based on reasonably good evidence, but that for some important parts of the model, the evidence is relatively poor. It is also evident that most of the model elements are value-neutral, and that conservative choices lead to an overall conservative model bias. The following variables should be prioritized for evaluating alternative hypotheses: tanker collision occurrence (V_{23}), tanker being striking or struck (V_{22}), size of the oil spill conditional to impact (V_{28}), tanker deadweight (V_6), impact angle (V_{19}) and impact location along tanker hull (V_{21}). Alternative hypotheses for following variables are less important, but can still be considered: tanker length (V_1), tanker mass (V_3), area in GoF (V_{15}), direction of tanker travel (V_{16}), models for oil spill mass (V_{25}, V_{26} and V_{27}) and the impact speeds of encountering vessel (V_{18}) and tanker (V_{20}). For these variables, alternative hypotheses are defined in the preceding sections. Their effects on the risk quantification are considered in Section 6.1.

5.5. Assumption effect assessment

An assumption assessment is performed in Table 7 using the method of Section 3.5, to assess how stable the risk quantification is with respect to assumptions made in the risk model construction. The use of the assumption assessment in the risk analysis can be understood by recalling the functions of the risk model as outlined in Section 2.3, especially the function to serve as an aid to thinking. Uncertainties and biases in the evidence (i.e. underlying the risk model construction) as well as uncertainties in the relation between the risk model and the space of possible outcomes are considered. Hence, the assumption effect assessment acts as a focal point to discuss the risk quantification of the first risk analysis stage, assisting the deliberative judgment leading to an uncertainty quantification in the second risk analysis stage,

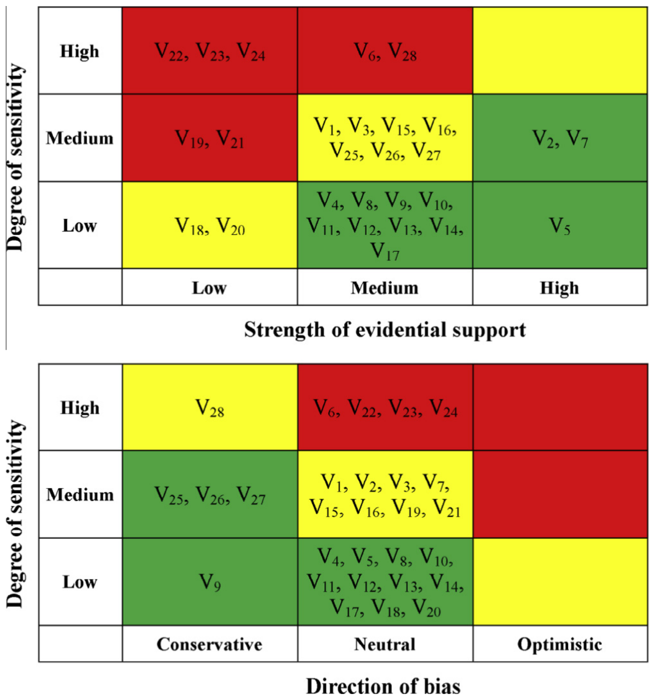

Fig. 5. Qualitative importance ranking of evidence uncertainty and bias for risk model variables, variable notations see Appendix A.

see Fig. 1. Some assumptions are rated and discussed below, to illustrate the rationale of assumption assessment scheme.

Assumption 1. The traffic composition is taken as 2011 baseline.

The data for describing the traffic configuration in the area is taken to provide good evidential support, see Table 4, but it dates from 2011 and may not be fully descriptive of future traffic conditions. More tanker traffic is expected in the Gulf of Finland, especially due to developments in Russian oil terminals (Brunila and Storgård, 2012). This may affect the number of large tankers operating in the area, which in turn affects the distribution of tanker lengths and deadweights. While it is quite certain that more large tankers will operate in the area, it is uncertain to what extent this will affect the size distribution: no studies on this issue are known. The deviation from this assumption is taken to be relatively modest, likely leading to an increase of the risk metric in the medium-higher range of the quantification.

Assumption 3. Operating conditions remain as in 2011 baseline.

The data assessment in Table 4 highlights that there is very little data for assigning a probability of collision occurrence. Uncertainty about this thus is high. Moreover, the historic accident data is in itself not very informative, as the context of the system operation has changed since the accidents occurred, and is likely to change also in the future. In recent years, there is more focus on tanker safety, with the tanker industry investing in extensive safety assessment programs (OCIMF, 2008). Moreover, various technological innovations such as on-line collision avoidance support (Mou et al., 2010), electronic communication of intended routes (Porathe et al., 2013) and enhanced navigation support services (Hänninen et al., 2014) can positively affect navigation safety. On the other hand, increase in traffic volume increases the number of vessel encounters and thus number of opportunities for a collision to result. Based on the above, it is considered likely that the collision occurrence is overestimated in the risk model.

Assumption 10. Only mechanical impact is considered as a consequence.

The risk model only considers oil outflow as a direct result of the mechanical damage to the hull. While there is little evidence available, some collision analyses account for the occurrence of fire and/or explosion caused by the collision (Klanac and Varsta, 2011) and for ship capsizing and sinking (IMO, 2008). This is not accounted for in the current risk model, but has the potential to lead to spill sizes beyond the maximum considered spill range, possibly leading to important deviations from the model-based risk picture of Fig. 6.

5.6. Risk quantification based on BN-model

The BN model results in a risk metric which combines the annual collision probability in the considered area with the potential spill sizes in case of an accident, through the BN-variable "probability of oil spill". The results of this analysis are shown in Fig. 6.

The influence of alternative hypotheses is considered by evaluating the model response for a set of test cases, shown in Appendix C. First, each hypothesis is altered in turn, keeping the other hypotheses at their base value, with the weights for the alternative hypotheses summarized in Appendix A. In these test cases, the most conservative (leading to highest risk quantification) and most optimistic (leading to lowest risk quantification) alternative hypotheses are applied. Finally, the extreme bounds for the model-based risk quantification are identified by setting all hypotheses first to their most conservative and then to their most optimistic

Table 6
Results of sensitivity analysis, variable notations see Appendix A.

Variable		$\max\lvert u_1\rvert$	Variable		$\max\lvert u_1\rvert$
V_{24}	Tanker as struck vessel	0.2811	V_{21}	Impact location along struck tanker hull	0.0050
V_{23}	Tanker collision occurrence	0.1456	V_{19}	Impact angle	0.0034
V_{28}	Oil spill mass in damaged tanks	0.0187	V_{16}	Direction of tanker travel	0.0015
V_6	Tanker deadweight	0.0098	V_{15}	Area in GoF	0.0009
V_{22}	Tanker striking or struck	0.0077	V_1	Tanker length	0.0006
V_7	Tanker loading condition	0.0051	V_2	Tanker width	0.0003

Table 7
Assessment of effect of assumptions on risk quantification.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Magnitude of deviation	L	L	L–M	L	L	L	L	L–M	L	M
Direction of deviation	+	±	–	±	±	±	+	–	–	+
Consequence range	M–H	L–H	L–H	L	L–H	L–H	L–H	L–H	L–H	L–B
Strength of justification	L	M	L–M	M–H	H	M–H	M	M–H	L	L

Label	Description	Appendix
A1	Traffic composition is taken as 2011 baseline	B.1.1
A2	Collision probability is equal for all encounters	B.2, B.4
A3	Operating conditions remain as in 2011 baseline	B.1.1, B.4
A4	Bunker capacity and arrangement are representative for all ships of size class	B.5.2
A5	Oil density is 0.9 tonne/m^3	B.5.2
A6	All cargo tanks and ballast tanks in cargo area have equal length	B.1.4
A7	Accidents are underreported at a 20% rate	B.4
A8	All oil is spilled from a breached tank	B.5.2
A9	All non-tanker encountering vessels are fully laden	B.1.2
A10	Only mechanical impact is considered as consequence	B.5

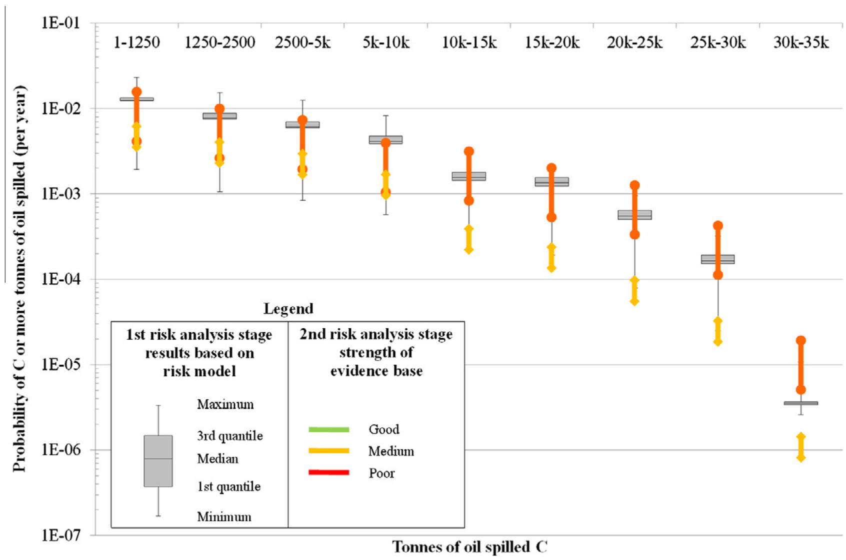

Fig. 6. Results from the risk model, alternative hypotheses as in Appendix C (risk analysis stage 1) and the uncertainty intervals (risk analysis stage 2), with evidence assessment as in Section 3.6.

values. The model response for these test cases is shown in Fig. 6 using summary statistics (median, quantiles, minima and maxima).

It is seen that the risk quantification is relatively stable in regards the general trend, despite the fact that alternative hypotheses can result in important changes in the risk metric, of just over an order of magnitude. This is in line with expectations, as various alternative hypotheses may lead to significantly different oil outflows, e.g. the impact angle (AH 5) and the damage extent model (AH 8 and AH 9).

However, the analysis indicates some general trends, which are stable despite the uncertainty. According to the model, spills in the range lower than 10,000 tonnes have an occurrence probability between ca. $4 \cdot 10^{-3}$ and $1 \cdot 10^{-2}$. Spills up to 20,000 tonnes have an occurrence probability of about $1 \cdot 10^{-3}$, whereas spills over 20,000 have a probability between $1 \cdot 10^{-4}$ and $3 \cdot 10^{-4}$. Spills over 30,000 tonnes are very unlikely in the area, with a model-based probability in the order of $3 \cdot 10^{-6}$. It is stressed that in the current framework, this quantification cannot be seen separate from the underlying evidence base, the qualitative evidence assessment and the assumption effect assessment.

6. Case study: risk analysis stage 2 and risk-informed decision making

6.1. Deliberative judgment and global evidence assessment

In the second risk analysis stage, the results of the model-based quantification (Fig. 6), the evidence assessment (Tables 4 and 5) and the assumption affect assessment (Table 7) are used to make a judgment in terms of degrees of belief of the occurrence of certain spill sizes. The uncertainty intervals are shown in Fig. 6, where a color code conveys information regarding the combined strength of the evidence for making the judgments. Two uncertainty interval series are shown. The intervals bounded by diamonds account for the model-based quantification, the evidence assessment and assumptions A1–A9, which is considered medium strength of evidence. The intervals bounded by circles additionally account for assumption A10, which concerns consequences beyond mechanical impact. As the evidence for these additional consequences is poor, the total evidence for this uncertainty interval is medium-poor, and the uncertainty interval is wider. It is seen that the first uncertainty interval leads to a lower oil spill risk than the risk model suggests, while the second uncertainty interval results in a higher oil spill risk than suggested by the risk model.

6.2. Application of risk analysis results in risk-informed decision making

The risk analysis results can be used to inform a decision. As outlined in Section 2.3, the results should be seen in a wider decision-making setting, where other issues such as costs and societal concerns are taken into account, e.g. in terms of the urgency of environmental protection. Hence, the oil spill risk from tanker collisions as resulting from the analysis can be used to update the knowledge underlying models for environmental risks from oil spills (Lecklin et al., 2011), for investigating the response fleet effectiveness in case of spills (Helle et al., 2011) and for determining the clean-up costs of spills in a sea area (Montewka et al., 2013). One immediate use of the risk analysis results relates to the required response capacity for oil spills in the Gulf of Finland. Presently, a capacity of 30,000 tonnes is available according to Jolma and Haapasaari (2014).

First, the risk analysis suggests that the occurrence of any accidental oil spill from tankers is unlikely, with an occurrence probability of any spill from tanker collisions of around 0.006. While this probability per se is not decisive in terms of the need for oil response preparedness, an indication of the occurrence of accidental spills can be relevant in the context of broader societal decision making for environmental protection. This can e.g. concern prioritizing investments in measures for other sources of oil spills, such as more frequently occurring operational spills (Hassler, 2011), measures for other types of pollution from maritime transportation or other sources of marine pollution in general.

Second, the analysis suggests that spills up to 30,000 tonnes may occur, but larger spills are very unlikely as far as only direct mechanical damage is considered. Smaller accidental spills, up to about 10,000 tonnes, are more likely than larger spills. While the evidence on which these results are based is not very strong in certain aspects (see Fig. 5), the consideration of alternative hypotheses indicates that the general trends are stable despite this uncertainty. The results could be used to argue that the current response capacity is sufficient, or even that the current capacity is over-conservative, as far as concerns tanker collisions. However, the analysis also highlights that considering further possible consequences such as fire, explosion and sinking due to collision could result in even larg-

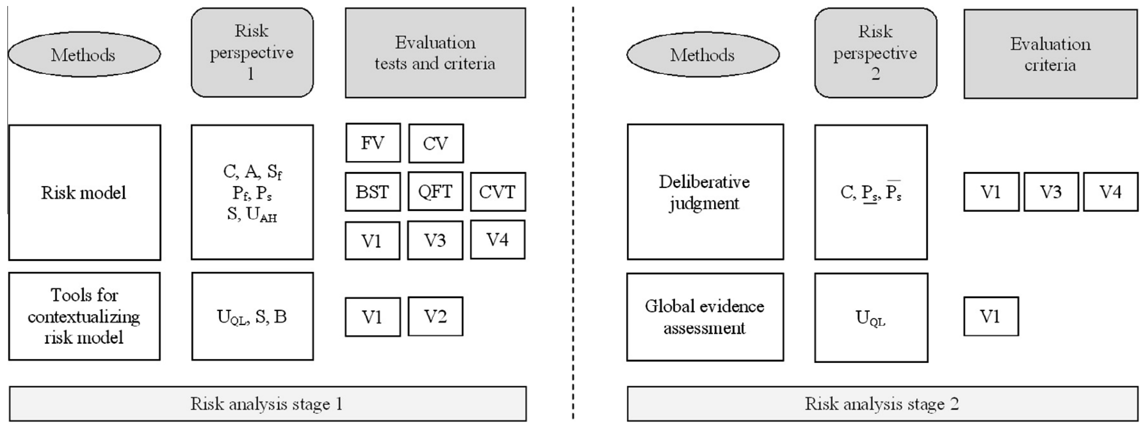

Fig. 7. Evaluation tests and criteria for the two risk analysis stages, in relation to the risk analysis methods and perspectives, see also Fig. 1.

er spills. Evidence for this is however scarce and further analysis is needed to reduce this uncertainty.

One of the reviewers of an earlier version of the manuscript suggested discussing the risk analysis results in the context of risk acceptance criteria (RAC). Here, it should be noted that for oil spill risk, there are no currently agreed bounding values for the ALARP[8] regions for maritime oil spill risk in a cumulative consequence probability plot, as exist e.g. for loss of life (Papanikolaou, 2009). Cost-effectiveness criteria for risk-reducing measures have been proposed (Vanem et al., 2008), but numerical values for oil spill costs are not agreed upon (Psaraftis, 2012). Moreover, these would not be applicable as the current risk model and analysis do not include risk-reducing measures (a limitation of the case study, not of the framework).

On a more fundamental level, it has been argued that defining RAC and evaluating that the calculated risk meets these criteria is not required for managing risk. First, defining criteria does not lead to more ethical risk management (Aven, 2007). Second, the introduction of pre-determined criteria may give a wrong focus, i.e. meeting these criteria rather than obtaining overall good and cost-effective solutions. Third, standard use of RAC presupposes that risk analyses can achieve an adequate precision level, which can be questioned depending on the strength of the knowledge base. Alternative decision making strategies exist, focusing on a broad assessment of risks, costs, public perceptions and other socio-economic concerns (Aven and Vinnem, 2005). The current framework, which acknowledges the weaknesses in the evidence base and the judgmental nature of risk analysis, therefore does not apply RAC. Rather, the analysis results are used in a broader risk evaluation in a managerial review and judgment, see e.g. Aven and Vinnem (2005).

7. Discussion: evaluation of the risk analysis

7.1. Evaluation: method and criteria

Validity concerns the question whether the analysis describes the specific concepts one intends to describe, for its intended use (Carmines and Zeller, 1979), whereas evaluation is a quality control process with the risk analysis as its object (Rosqvist and Tuominen, 2004). Two aspects are considered. First, the plausibility of the risk model as a tool for serving its envisaged functions in the risk analysis is assessed. As introduced in Section 2.3, the model functions: (i) to convey an argumentation based on available evidence, (ii) to provide a basis for communication and (iii) to serve as an aid to thinking. Its plausibility is evaluated using model-construct and risk-theoretical validity tests. Second, the risk analysis *qua* risk analysis is evaluated using relevant criteria. The evaluation tests and criteria are shown in Fig. 7 in relation to the two defined stages for risk analysis. These criteria and tests are briefly further outlined below: the reader is referred to the cited publications for more elaborate discussions.

7.1.1. Evaluation of the risk model

As a reflection of a mind construct addressing possible consequences which may or may not occur, a direct comparison between the risk model results and observations from the described system is not possible. In the considered area, no accidental oil spills from tankers have been reported during the period 1998–2014 (HELCOM, 2014), so a comparison with historic data is likewise inconclusive. However, evaluation can be understood in a wider sense than a comparison with observed data, by inspecting the model *qua* model. Such approaches are widely used in social science research (Trochim and Donnely, 2008), system dynamics modeling (Forrester and Senge, 1980) and for expert-based Bayesian Network modeling (Pitchforth and Mengersen, 2013). These model-related tests are only of interest in the first risk analysis stage, see Fig. 7.

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 face and content validity. Face validity (FV) is a subjective, heuristic interpretation of whether the model is an appropriate operationalization of the construct. Content validity (CV) 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. A behavior sensitivity test (BST) is used to assess to which model elements the results are sensitive. The parameter sensitivity of a BN can be calculated as in Section 3.1.3, 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 how the system is believed to respond under these conditions. In a concurrent validity test (CVT), 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.[9]

[8] ALARP: As Low As Reasonably Practicable.

[9] More model evaluation tests have been proposed in the literature then the ones retained here, e.g. a dimensional consistency tests, boundary adequacy tests and structure verification tests (Forrester and Senge, 1980). Which tests are considered largely depends on the type of developed model. For the purposes of this paper, a limited number of relatively straightforward tests is retained.

The validity tests do not "prove" that the model results are correct, but only indicate the extent to which the model is a plausible representation of the object of inquiry, serving functions (i) and (ii) as outlined above. This relates to adopted understanding of risk and the adopted risk perspective, where no reference is made to an underlying "true" risk, see Section 2.2. The model should be plausible enough to serve as a basis for further reflections, leading to deliberative judgments in the second risk analysis stage, see Fig. 7.

7.1.2. Evaluation of the risk analysis

Following criteria are considered in the current framework, based on work by Aven and Heide (2009) and Rosqvist and Tuominen (2004):

- V1: the degree to which the uncertainty assessments are complete.
- V2: the degree to which the bias assessments are complete.
- V3: the degree to which the assigned subjective probabilities adequately describe the assessor's uncertainties of the unknown quantities.
- V4: the degree to which the analysis addresses the right quantities.

As illustrated in Fig. 7, V1 and V2 concern the uncertainty and bias related to the limitations of the risk model to cover the scope of the possible outcome space, as well as the uncertainty and bias concerning the evidence for assessing probabilities in the model construction. V1 also concerns the uncertainty judgments in the second risk analysis stage, and the uncertainty in the evidence for making these uncertainty judgments (i.e. the global evidence resulting from the first risk analysis stage). V3 is relevant in both risk analysis stages, and concerns whether appropriate elicitation principles and procedures are followed to elicit the probability judgments. An elaborate discussion on this criterion is outside our current scope, for more details see Ayyub (2001) and O'Hagan et al. (2006). V4 addresses the question whether the analysis focuses on fictional quantities (parameters of a model) or on observables (events, consequences of observable aspects of a system). This last criterion thus relates to the interpretability of the risk analysis results, i.e. how easily the presented numbers and information can be given a meaning, see also Aven (2011a).

It can be noted that the risk model evaluation criteria can be related to the evaluation of the risk analysis, especially to V1, the completeness of the uncertainty assessments. For example, an evaluation of the content validity can assist in evaluating the adequacy of the evidence base and the underlying assumptions, as in Sections 3.3 and 3.5.

7.2. Application of evaluation criteria to the risk analysis

7.2.1. Risk model evaluation tests

In terms of face validity, it can be found that the risk model is an adequate reflection of ship–ship collision oil spill risk. The sequence encounter-impact-hull breach is a logical flow of events for an oil spill to occur, and the elements describing these situations seem reasonable. Encounters occur between ships of certain types and dimensions in certain locations and encounter types, and at certain speeds and loading conditions. The situation at impact is somehow related to the encounter conditions, and aspects such as which is the striking and struck vessel, the impact location, speeds and impact angle are known to affect the damage size. The vessel size is related to the size of the oil tanks, and hence the possible spill sizes. Content validity can similarly be established, by more carefully inspecting the adequacy of the elements of the model in relation to knowledge about the system. For reasons of brevity, this is not elaborated upon here. Considering the comment made in Section 7.1.2 concerning the relation between CV and V1, the reader is referred to Sections 5.3 and 5.5, where the evidence and assumption assessments are performed.

The behavior sensitivity test is performed using the methodology described in Section 3.1.3, with results shown in Table 6. It is seen that the model output is mainly sensitive to changes in the parameters of the variables "Tanker as struck vessel", "Tanker collision occurrence", "Oil spill mass in damaged tanks", "Tanker deadweight", "Tanker striking or struck", "Tanker loading condition", "Impact location along struck tanker hull" and "Impact angle". It is found that the sensitivity of the model results to these variables is plausible: the oil outflow is clearly sensitive to the occurrence of a collision, whether the tanker is striking or struck, on the size of the tanker and its loading condition and on the impact conditions which govern how many tanks are breached.

The qualitative features test is performed for a number of test conditions, by selecting certain states of the BN-variables as inputs, and by inspecting the corresponding model response. This is illustrated in Table 8, showing the test settings, the expected value of the oil spill risk measure and a short assessment of the plausibility of the results. It is seen that the results indicate that the BN-model qualitatively follows the expected response.

Concurrent validity can be tested by inspecting the structure and content of models for similar problems to the one developed. In Fig. 8, a number of models for oil spill risk is briefly reviewed, focusing on structure (the logical sequence of events leading to

Table 8
Results from the qualitative features test for the BN-model of Fig. 4.

Variable	State	*E* (C/year)	Plausibility assessment
Baseline	Appendix A	29.8	–
Tanker length	100–120	7.7	Small tankers carry less oil than large ones, implying that a collision leads to smaller spills
	240–260	66.1	
Tanker encounter speed	4–8	18.2	Lower encounter speeds imply lower impact speeds and collision energy, smaller hull damages and
	12–16	30.2	smaller spills
Tanker loading condition	Laden	55.9	Laden tankers lead to bigger spills from cargo tanks. In ballast condition, only bunker tanks can lead
	Ballast	1.3	to a spill
Impact angle	0–36	7.3	Oblique angles may inflict insignificant damage if vessels grind alongside, perpendicular impact
	72–108	39.2	leads to damage
Impact location	0–20	12.9	In aft ship, impact does not lead to spill or to small bunker tank spills. In midship, spills from cargo
	40–60	39.3	tanks are larger
Tanker striking or struck	Striking	0	If the tanker is the striking vessel, no spills occur. Otherwise, spills occur
	Struck	45.9	
Damage extent model	M1	31.6	By construction, M2 leads to smaller damages than M1 as more stringent limits are imposed on
	M2	29.6	bulkhead breach

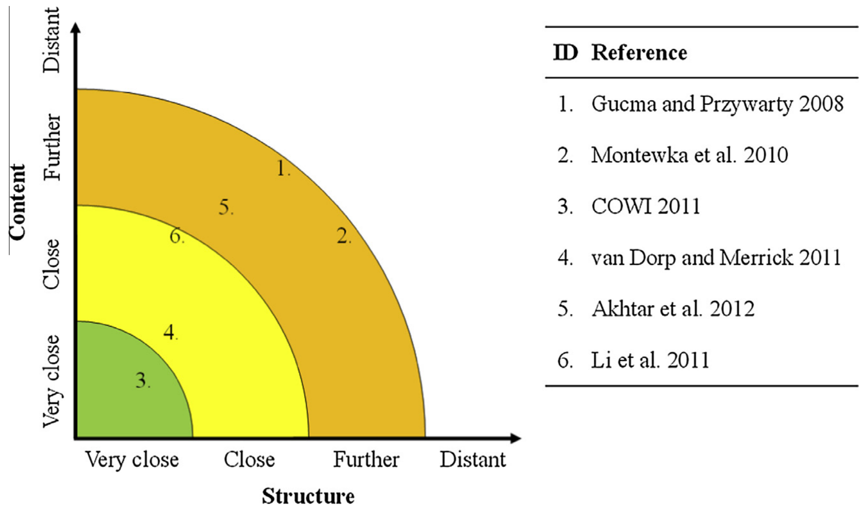

Fig. 8. Results from concurrent validity test for model structure and content.

Table 9
Comparison of concurrent validity (structure and content) for selected models.

	Gucma and Przywarty (2008)	van Dorp and Merrick (2011)	Proposed risk model
Encounter	AIS data Traffic simulation model	AIS and environmental data Traffic simulation model	AIS data Encounter detection method Alternative hypotheses
Collision occurrence	Based on accident data	Expert-based model	Based on accident data Accounts for underreporting Alternative hypotheses
Impact	*not explicitly modeled*	Assumed relations between encounter and impact	Models linking encounter to impact Alternative hypotheses
Hull damage	*not explicitly modeled*	Meta-model based on coupled collision mechanics model Two tanker designs considered	Meta-model based on coupled collision mechanics model Multiple tanker designs based on tanker data Alternative hypotheses
Oil outflow	Based on accident data	Cargo tank layout of two ships Bunker tank data	Cargo tank layout multiple ships Bunker tank data

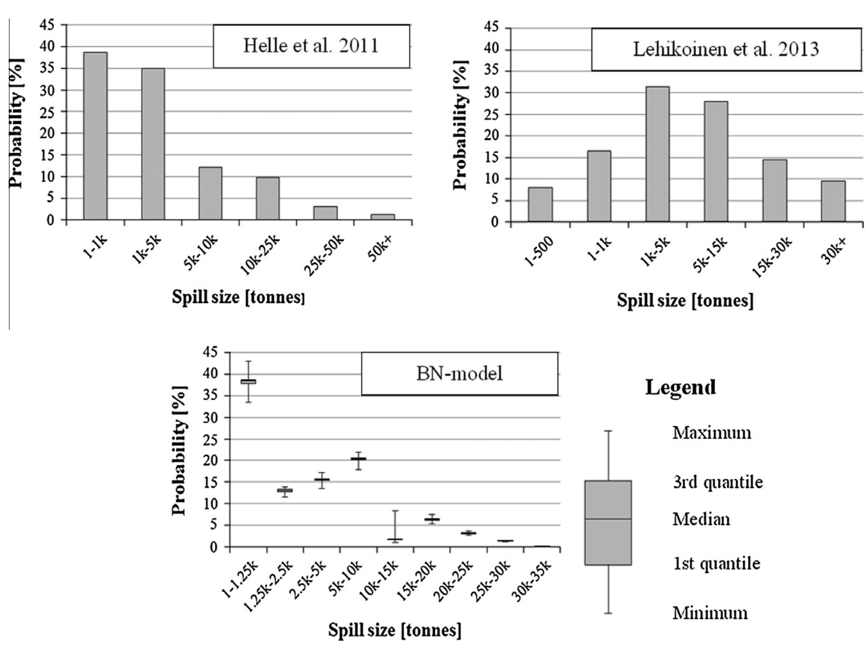

Fig. 9. Results from concurrent validity test for model structure and content; for BN-model, the summary statistics are calculated for model output for test cases of Appendix C.

Table 10
Advantages and drawbacks of the risk analysis tools.

Tool	Benefits	Drawbacks
Bayesian Network model Sections 3.1 and 5.1	• Provides quantitative risk picture • Integrates various sources and qualities of evidence (-data, models, judgments) • Explicit treatment of uncertainty, including alternative hypotheses (AHs) • Computationally efficient • Allows sensitivity analysis • Focus on observable system aspects • Relatively accessible to non-experts	• Does not distinguish well-supported from poorly evidenced probabilities • Development of expert-based BNs is resource-intensive • Stable causal dependencies between system parts are to be identified or assumed • Discretization leads to information loss • No support for modeling system feedback
Evidence assessment Sections 3.3 and 5.3	• Increases transparency of analysis • Easy to perform and use • Shows strengths, weaknesses and biases • Acts as a guide for improving analysis	• Relatively crude • Quantity of ratings can obscure their relative importance
Evidence importance ranking and selection of alternative hypotheses Sections 3.4 and 5.4	• Visual representation of evidence strength and direction of bias • Easy to perform and use • Highlights important model variables, for which AHs can be applied	• Relatively crude • Importance is relative to risk model, thus constrained by underlying choices
Alternative hypothesis approach Sections 3.2 and 5.6	• Shows effect of conflicting evidence • Insight in stability of risk quantification	• Utility constrained by quality of underlying evidence • Relatively resource-intensive • Infeasible to consider all alternatives
Assumption effect assessment Sections 3.5 and 5.5	• Provides insight in importance of assumptions for considered problem • Serves as a basis for discussion and further evidence gathering • Easy to perform and use • No constraints on data or models	• Relatively crude • Utility constrained by understanding of analyst team
Global evidence assessment Sections 3.6 and 6.1	• Provides a global insight in the combined strength of evidence for the risk quantification • Easy to perform and use	• Relatively crude

spills) and content of the models (which elements are considered in analyzing the consequences). For reasons of brevity, not all models are analyzed in detail in the comparison. Two models are briefly considered.

The models by Gucma and Przywarty (2008) and by van Dorp and Merrick (2011) follow implicitly or explicitly the sequence of events underlying the model presented in Fig. 4: encounter, impact, hull damage and oil outflow. A brief comparison is made in Table 9. It can be concluded that the presented model has significant similarities to some other models in the literature, both in structure and content, while more thoroughly accounting for uncertainties through considering alternative hypotheses. As the focus of the presented model is the magnitude of the consequences, the impact, hull damage and tanker capacity is modeled in more detail than in most other oil spill risk models, whereas the level of detail in the collision occurrence is comparable to most other oil spill risk models (Fig. 9).

A final concurrent validity check can be performed by comparing the outcome of the model with other information regarding oil spills. As mentioned in Section 7.1.1, no accident data is available in the study area, so a comparison with spill data cannot be performed. Moreover, as the models mentioned above are applied to different sea areas, these cannot be used as a comparison either. However, other models have been proposed for the Gulf of Finland, where oil spill sizes where included. The distributions in Helle et al. (2011) and Lehikoinen et al. (2013) cover both collision and grounding accidents, and show the probability of a spill of a certain size in case an oil spill occurs. The distributions are respectively based on simple outflow models (IMO, 2003; Montewka et al., 2010b), which only account for oil outflow due to mechanical damage and do not take local traffic conditions into account. The results from the current model differ from the earlier results mainly by the lower probabilities for the larger spill sizes, but a comparison is difficult as the analysis scope is not identical. As the simple outflow models (IMO, 2003; Montewka et al., 2010b) have some serious limitations to maritime transportation risk analysis as argued by van de Wiel and van Dorp (2011), it is concluded from these tests that the proposed model is a more accurate reflection for the probability of different oil spill sizes in the Gulf of Finland then achieved by the earlier models.

7.2.2. Evaluating the risk analysis

At the first stage of risk analysis, the criteria V1 and V2 are addressed by performing the evidence assessment (Tables 4 and 5), the alternative hypotheses in the risk model (Figs. 4 and 6) and by the assumption effect assessment (Table 7). As found also in Aven and Heide (2009), there is no guarantee that all uncertainty is addressed. However, the assessments increase the transparency of the evidence base, and indicate the strength of the argument put forward by the risk model. Criterion V4 is met, as the probability assignments focus on observable quantities of the maritime transportation system: the situational factors and events of Fig. 4 are observables in the maritime transportation system and in collision accidents.

At the second stage of risk analysis, the uncertainty assessment (V1) is performed by assessing the global strength of evidence, and by assessing degrees of beliefs over the outcome space based on the global evidence, see Fig. 6. The analysis also focuses on observable quantities (V4), namely the spill sizes in collision accidents with tankers.

Criterion V3 is difficult to verify: the subjective probabilities used at the first stage (e.g. for BN-variables V_{18}, V_{19} and V_{20}, see Table A1 in Appendix A) and probability intervals at the second stage are assessor's judgments. The principles and procedures provided in Aven and Heide (2009) are followed to the extent possible in stage 2, but as the subjective probabilities in the first stage are obtained from the literature, no information is available on how these are assessed.

8. Discussion: reflection on the methods for contextualizing risk

In Sections 2 and 3, various tools have been introduced to quantify and highlight uncertainty, and to convey qualitative information beyond the risk model. These tools have been applied in Sections 5 and 6. Some reflections on the benefits and drawbacks of these tools are given in Table 10. For further discussions on the advantages and challenges of Bayesian Networks, see Hänninen (2014) and Uusitalo (2007).

The use of the various methods for contextualizing the risk quantification in the first risk analysis stage depends on the decision context. For instance, in a particular application, it can be decided not to perform the evidence importance ranking and the alternative hypothesis approach, because this is relatively resource intensive. The tools for providing insight in the strength of the evidence and the effect of assumptions are less demanding, and can also be used to contextualize the risk quantification.

9. Conclusion

In this paper, a framework for risk analysis for maritime transportation systems is proposed, following a two-stage procedure. At the first stage, directed at an expert-review, a Bayesian Network model is constructed, providing a basis for communicating evidence and putting forward an argumentation based on this evidence. This is supplemented with broad evidence and assumption effect assessments, functioning as an aid to thinking beyond the model, qualifying the model-based argument. Sensitive model elements are identified, and if possible, alternative hypotheses are applied to quantify the uncertainty within the BN model. The second stage, directed at decision makers, consists of a deliberative judgment through assessing subjective probabilities. These are informed by the results of the first analysis stage. A global qualitative evidence assessment conveys information regarding the combined strength of the evidence. Various tools have been proposed to highlight uncertainties and biases beyond the model, and their merits have been discussed.

The framework has been applied to a case study of spills from collisions with oil tankers in a given sea area, focusing on the possibility of occurrence of certain spill sizes. While the model is applied to the Gulf of Finland, the approach can be adapted to any sea area, providing a proactive method for supporting decision making concerning oil spill risk. The evaluation tests indicate that the model is a plausible representation of the oil spill risk, accounting for many factors found relevant also in other analyses. Model behavior and qualitative features tests also indicate the model's plausibility. The evidence assessment shows that the model provides adequate decision support e.g. for spill capacity planning, but that further improvements can be made by more accurate modeling of the damage size and oil outflow, and by analysis of further consequences initiated by the collision event.

The presented framework applies risk analysis only as decision support, not to make firm recommendations or "optimized" solutions about what should be done. In this, the framework differentiates itself from some other maritime risk analysis frameworks and applications, by more explicitly and elaborately communicating and reflecting on the limitations of the risk quantification. It is hoped that the tools for treating uncertainty and bias can be further developed and applied also in other maritime risk analysis, especially because earlier research has highlighted lack of uncertainty treatment in the application area, and lack of tools for assessing uncertainty.

List of symbols and abbreviations

$\boldsymbol{A}$	arcs in BN
$\boldsymbol{AH}$	set of alternative hypotheses in BN
AIS	Automatic Identification System
ALARP	As Low As Reasonably Practicable
$\boldsymbol{AR}$	risk metrics in BN
B	qualitative bias assessment/beyond maximum
BK	background knowledge
BN	Bayesian Network
BST	behavior sensitivity test
C	consequence/conservative
COLREGs	collision regulations
C_i	volumetric coefficient
CPT	conditional probability table
CT	number of center tanks
CV	content validity
CVT	concurrent validity test
DAG	directed acyclic graph
E	event
$\boldsymbol{E}$	set of events in BN
EB_i	alternative i in evidence base
Ev_{QL}	qualitative evidence assessment scheme
FSA	Formal Safety Assessment
FV	face validity
$G(\boldsymbol{V},\boldsymbol{A})$	graphical component of BN
GOFREP	Gulf of Finland reporting system
H	high
IMO	International Maritime Organization
JDP	joint probability distribution
$\max\lvert u_1\rvert$	maximum sensitivity value of BN-variable
L	low
L_A	distance from aft perpendicular to closest transverse cargo tank bulkhead
L_F	distance from fore perpendicular to closest transverse cargo tank bulkhead
L_T	cargo tank length
LBH	longitudinal bulkhead positions
LNG	Liquid Natural Gas
M	medium
M_{bal}	ship mass in ballast condition
M_{lad}	ship mass in laden condition
MMSI	maritime mobile service identity
N	neutral
NBT	number of ballast tanks
NCT	number of cargo tanks
O	optimistic
$\boldsymbol{P}$	probabilistic component of BN
$\boldsymbol{Pa}(V_i)$	parent nodes of BN-variable V_i
p_i	subjective weight, expressing credibility of model i to reflect underlying mechanism
P_f	frequentist probability
P_s	subjective probability
$[\underline{P_s}, \overline{P_s}]$	probability interval lower and upper bound
QFT	qualitative features test
RAC	risk acceptance criteria
RM	risk metric
S	sensitivity analysis
S_f	situational factor
$\boldsymbol{S_f}$	set of situational factors in BN
$\boldsymbol{S_f^E}$	BN-variables concerning encounter situation

(continued on next page)

$\boldsymbol{S_f^I}$	BN-variables concerning impact situation
ST	number of side tanks
TBH	transverse bulkhead positions
TSS	Traffic Separation Scheme
TT	tank configuration type
U_{AH}	quantitative uncertainty assessment using alternative hypotheses
U_{QL}	qualitative assessment of uncertainty
$\boldsymbol{V}$	BN-variables
V_{BT}	volume of ballast tank
V_{CT}	volume of cargo tank
V_i	BN-variable/tank volume/risk analysis evaluation criterion
VTS	Vessel Traffic Service
y_L	collision damage length
y_T	collision penetration depth
θ	collision damage direction
ρ_c	cargo oil density
ρ_{sw}	sea water density
+	increase
±	increase/decrease
–	decrease

Acknowledgements

This work presented in this paper is carried out within the WINOIL project in association with the Kotka Maritime Research Centre. This project is co-funded by the European Union, the Russian Federation and the Republic of Finland. The contributions by the second author are supported by the EU-funded project FAROS (Grant No. 314817), the FP7 program. The financial support is acknowledged. The opinions expressed are those of the authors and should not be construed to represent the views of the project consortia. The presented Bayesian Network model has been developed using the GeNie modeling environment developed at the Decision Systems Laboratory, University of Pittsburgh, available from http://genie.sis.pitt.edu. The authors are grateful to three anonymous reviewers whose thoughtful comments have helped to improve an earlier version of this paper.

Appendix A. Notations and discretization of BN-model variables

See Table A1.

Appendix B. Evidence underlying the Bayesian Network model

In this appendix, the evidence for the construction and parameterization of the BN-model shown in Section 5.1 is briefly presented, for the case study of the Gulf of Finland, see Section 4. Brevity is aimed at, but it is attempted to describe the data processing, to facilitate potential application of the model to other sea areas.

B.1. Evidence related to vessel characteristics

B.1.1. Data related to vessel traffic in the sea area

Data related to the main vessel characteristics operating in the area are obtained from the Automatic Identification System (AIS) for the period 2011–07 to 2011–11, with data fields as shown in Table B1. The data is further processed using an encounter detection model, see Section B.2.

Concerning the ship types of encountering vessels, the AIS data only specifies the crude categories cargo ship, passenger ship and tanker. AIS data analysis leads to the probabilities for the variable "encountering vessel type (AIS)", see Section B.2. For evaluating the damage extent, the half bow entrance angle of the encountering vessels is required, see Section B.5. Data for this angle is available for container ships, bulk carriers and general cargo vessels. Based on a traffic analysis by Nyman et al. (2010), a proportion between these cargo vessels is obtained as shown in Table B2. This is used in the CPT for the variable "encountering vessel type (detailed)". For the tankers, only tankers with AIS classification "tankers – all types" are considered. Other AIS ship classifications include chemical tankers and Liquid Natural Gas (LNG) tankers, which are not further considered here.

B.1.2. Data related to encountering vessels

The model for evaluating the damage extent and oil outflow (Section B.5) requires the vessel mass and the bow entrance angle η. The latter parameter is obtained for the considered vessel types from an analysis by Brown (2002), see Table B3. The mass of the encountering vessels M_{EV} is conditional to the vessel type, size and loading condition. The mass of fully laden general cargo, bulk carrier, container and passenger vessels is derived from data-based regression models presented by Brown (2002). The models, having a statistical fit with R^2-values around 0.98, have regression coefficients as in Table B3 and a functional form as follows, with L the ship length:

$$M_{EV} = \sqrt[a]{\frac{L}{c}} \tag{B1}$$

Except for tankers, there is no information available concerning the loading condition of the vessels operating in the area. Therefore, the conservative assumption is made that all are fully laden. For tankers, more detailed data and models are available, see Section B.1.4.

B.1.3. Data related to cargo of oil tankers

For oil tankers, data concerning the main dimensions, mass, deadweight and tank configuration is available from a ship database (IHS Maritime, 2013), for 410 oil tankers operating at least twice in the area during the period 2011–07 to 2011–10. It is assumed that these vessels are representative to the entire oil tanker fleet in the area. This is shown in Fig. B1, where L_2, B_2 and D_2 are respectively the tanker length, width and depth. TT signifies the tank configuration type. TT1 is a double hull (DH) tanker without longitudinal bulkhead, TT2 with one longitudinal bulkhead and TT3 with two longitudinal bulkheads. ST and CT signify the number of side and center tanks, respectively.

The loading conditions of the tankers directly affect the likelihood of a spill and depend on the tanker type, tanker size, location in the sea area and direction of travel (COWI, 2011). As the available AIS data does not contain data for specific tanker types, results of a traffic flow analysis by Nyman et al. (2010) are used to assess the probability of the tankers being oil product or crude oil tankers, see Table B4.

The quantification of the loading condition is based on a detailed analysis of goods transported in the Baltic Sea, reported by COWI (2011). The probabilities assessed based on this information are given in Table B5.

The mass of the tanker is derived directly from the ship database, see Fig. B1, if the vessel is fully laden. If the vessel is in ballast condition, the tanker mass is determined based on the ship database and a model for cargo and ballast tank configuration, see Section B.1.4.

B.1.4. Model for cargo and ballast tank configuration and tanker mass in ballast condition

The cargo tank dimensions are determined based on a model proposed by Smailys and Česnauskis (2006). The main parameters

Table A1
Description of variables in the model.

Factor group Variable		Symbol	Discretization		Ref.
Encounter situation – tankers ($\boldsymbol{S_f^E}$)					
V_1	Tanker length	L_2	{L80, 80–100, 100–120, 120–140, 140–160, 160–180, 180–200, 200–220, 220–240, 240–260, M260} [m]		B.1
V_2	Tanker width	B_2	{L15, 15–20, 20–25, 25–30, 30–32.5, 32, 5–40, 40–45, 45–50} [m]		B.1
V_3	Tanker mass	M_T	{L10, 10–15, 15–20, 20–30, 30–40, 40–60, 60–80, 80–100, 100–120, 140–190} [ktonne]		B.1
V_4	Tanker encounter speed	V_2^E	{0–4, 4–8, 8–12, 12–16} [kn]		B.1
V_5	Tanker type	–	{Oil product, crude oil}		B.1
V_6	Tanker deadweight	–	{L10, 10–20, 20–30, 30–40, 40–80, 80–120, 120–160} [ktonne]		B.1
V_7	Tanker loading condition	–	{Laden, Ballast}		B.1
Encounter situation – encountering vessel ($\boldsymbol{S_f^E}$)					
V_8	Encountering vessel length	L_1	{L80, 80–100, 100–120, 120–140, 140–160, 160–180, 180–200, 200–220, 220–240, 240–260, M260} [m]		B.1
V_9	Encountering vessel mass	M_{EV}	{L10, 10–20, 20–40, 40–80, 80–180} [tonne]		B.1
V_{10}	Encountering vessel type (AIS)	–	{Tanker, Passenger vessel, Cargo vessel}		B.1
V_{11}	Encountering vessel type (detailed)	–	{Tanker, Passenger vessel, Bulk carrier, General cargo, Container vessel}		B.1
V_{12}	Encountering vessel encounter speed	V_2^E	{0–8, 8–12, 12–16, M16} [kn]		B.1
V_{13}	Encountering vessel loading condition	LC_{EV}	{Laden, Ballast}		B.1
V_{14}	Encountering vessel bow entrance angle	η	{L34, 34–40, M40} [°]		B.1
Encounter situation – context ($\boldsymbol{S_f^E}$)					
V_{15}	Area in Gulf of Finland	–	{Russian sector, Finnish GOFREP, Estonian GOFREP}		B.2
V_{16}	Direction of tanker travel	–	{Inbound, outbound}		B.2
V_{17}	Encounter type	–	{Overtaking, Meeting, Crossing}		B.2
Impact situation ($\boldsymbol{S_f^I}$)					
V_{18}	Encountering vessel impact speed	V_1^I	{0–8, 8–12, 12–16, M16} [kn]		B.3
V_{19}	Impact angle	φ^I	{0–36, 36–72, 72–108, 108–144, 144–180} [°]		B.3
V_{20}	Tanker impact speed	V_2^I	{0–4, 4–8, 8–12, 12–16} [kn]		B.3
V_{21}	Impact location along struck tanker hull	l	{0–20, 20–40, 40–60, 60–80, 80–100} [from stern]		B.3
V_{22}	Tanker striking or struck	–	{Striking, Struck}		B.3
V_{23}	Tanker collision occurrence	–	{Collision, No collision}		B.4
V_{24}	Tanker as struck vessel	–	{Tanker struck, Tanker not struck}		B.4
Consequences – oil spills from damaged compartment (C)					
V_{25}	Oil spill mass – AH 8 – Model 1	C_{M1}	{0, 1–1250, 1250–2500, 2500–5 k, 5 k–10 k, 10 k–15 k, 15 k–20 k, 20 k–25 k, 25 k–30 k, 30 k–35 k} [tonne]		B.5
V_{26}	Oil spill mass – AH 8 – Model 2	C_{M2}	{0, 1–1250, 1250–2500, 2500–5 k, 5 k–10 k, 10 k–15 k, 15 k–20 k, 20 k–25 k, 25 k–30 k, 30 k–35 k} [tonne]		B.5
V_{27}	Oil spill mass – AH 8 weighed	C_{AH8}	{0, 1–1250, 1250–2500, 2500–5 k, 5 k–10 k, 10 k–15 k, 15 k–20 k, 20 k–25 k, 25 k–30 k, 30 k–35 k} [tonne]		B.5
V_{28}	Oil spill mass in damaged tanks	C_{AH9}	{0, 1–1250, 1250–2500, 2500–5 k, 5 k–10 k, 10 k–15 k, 15 k–20 k, 20 k–25 k, 25 k–30 k, 30 k–35 k} [tonne]		B.5
Alternative hypotheses (***AH***)			Alternatives	Weights	
AH concerning encounter detection		AH 1	{M1, M2}	{0.5, 0.5}	B.2
AH encountering vessel impact speed		AH 2	{M1, M2, M3, M4}	{0.1, 0.1, 0.1, 0.7}	B.3
AH tanker impact speed		AH 3	{M1, M2, M3, M4}	{0.1, 0.1, 0.1, 0.7}	B.3
AH impact location along struck tanker hull		AH 4	{M1, M2}	{0.25, 0.75}	B.3
AH impact angle		AH 5	{M1, M2, M3, M4, M5, M6}	{0.05, 0.2, 0.1, 0.2, 0.05, 0.4}	B.3
AH tanker striking or struck		AH 6	{M1, M2}	{0.5, 0.5}	B.3
AH tanker collision probability		AH 7	{M1, M2}	{0.5, 0.5}	B.4
AH damage extent and oil outflow		AH 8	{M1, M2}	{0.1, 0.9}	B.5
AH damage extent oblique impact angles		AH 9	{M1, M2}	{0.2, 0.8}	B.5
Risk metric (***RM***)					
Probability of oil spill		–	{0, 1–1250, 1250–2500, 2500–5 k, 5 k–10 k, 10 k–15 k, 15 k–20 k, 20 k–25 k, 25 k–30 k, 30 k–35 k} [P (tonne/year)]		6.1

for the determination of the tank volumes and the location of the transverse and longitudinal bulkheads are shown in Fig. B2. L_F and L_A are the distances from the fore and aft perpendicular to the respective closest transverse cargo tank bulkheads. The positions of the transverse bulkheads TBH are determined based on L_A, L_F, the tanker length L and the number of cargo tanks. Equal cargo tank lengths L_T are assumed. The positions of the longitudinal bulkheads LBH are determined based on the tanker width B_2, the width of the double hull w and number of longitudinal bulkheads, i.e. the tank configuration type. For TT2, the bulkhead is at the center line and for TT3, the two bulkheads are at a distance $w + \frac{2}{7}(B_2 - 2w)$ from the ship side.

The volume V_i of a given tank is determined as follows, with notations as in Fig. B2:

$$V_i = C_i B_T L_T D_T \tag{B2}$$

Table B1
AIS data fields available for the presented model.

Data field	Unit	Explanation
MMSI number	–	A 9-digit code uniquely identifying a vessel
Time stamp	s	Time at which the message is recorded [YYYY]–[MM]–[DD] [hh]:[mm]:[ss]
Position		Longitude and latitude of transmitted message, in WGS-84 coordinate system
Ship type	–	A 2-digit code identifying the type of vessel, see USCG (2012)
Ship length and width	m	Dimensions from bow to stern and side to side, see USCG (2012)
Ship speed	kn	Speed over ground
Ship course	°	Course over ground

C_i is a volumetric coefficient, accounting for the actual shape of the tank in comparison with a rectangular prism, assuming a cargo loading level of 98%. Values for C_i depend on the tank configuration type, the location of the tank and the size of the vessel (Smailys and Česnauskis, 2006).

The model also enables an evaluation of the ballast water volume, from which the tanker mass in ballast condition can be derived, as follows:

$$M_{T,bal} = M_{T,lad} - \rho_c \sum_{j=1}^{NCT} V_{CT,j} + \rho_{sw} \sum_{k=1}^{NBT} V_{BT,k} \tag{B3}$$

where M_{bal} is the ship mass in ballast condition, M_{lad} the mass in laden condition as available in the data shown in Fig. B1, ρ_c the density of the cargo oil, taken as 900 kg/m^3 and ρ_{sw} the density of sea water, taken as 1025 kg/m^3, NCT and NBT the total number of cargo and ballast tanks, $V_{CT,j}$ the volume of the j-th cargo tank, obtained from Eq. (B2), and $V_{BT,k}$ the volume of the k-th ballast tank, determined as follows:

$$V_{BT}^{dh} = wD_T L_T \tag{B4}$$

$$V_{BT}^{db} = h\left(\frac{B - \sum_i^m B_{T,i}}{2}\right) L_T \tag{B5}$$

$$V_{BT}^{fp} = 0.4D^3 \tag{B6}$$

with V_{BT}^{dh} a ballast tank in the double hull, V_{BT}^{db} a ballast tank in the double bottom, V_{BT}^{fp} the forepeak ballast tank, calculated as in Okumoto et al. (2009) and m the number of cargo tanks in a transverse section. The other notations are as applied above.

B.1.5. Data and information related to bunker tank size and position

The model for the oil outflow of Section B.5 includes the bunker oil. Limited data regarding the bunker tank volumes as well as information of some common bunker tank layouts is available from McAllister et al. (2003). It is assumed that bunker tanks are fully laden. The volume of the bunker tanks for different vessel categories, defined according to Evangelista (2002), is given in Table B6, along with assessed probabilities of the various tank configurations. The alternative tank configurations are shown in Fig. B3.

Table B2
Probabilities of cargo vessel types in various sea areas, based on Nyman et al. (2010).

	Container	Bulk carrier	General cargo
Finnish GOFREP	0.42	0.08	0.5
Estonian GOFREP	0.42	0.08	0.5
Russian national VTS	0.44	0.07	0.49

Table B3
Regression coefficients for encountering vessel mass and bow entrance angle η, based on Brown (2002).

Ship type	c	a	η
Bulk carrier	6.6	0.332	40
General cargo	6.93	0.325	40
Container ship	5.49	0.353	34
Passenger ship	8.22	0.299	34
Tanker	N/A	N/A	76

B.2. Evidence for situational factors at vessel encounters

For deriving probabilities for various factors related to the encounter situation, a model is used to detect encounters between tankers and encountering vessels in the AIS data described in Section B.1.1. Probability distributions are constructed for the type, length and speed of encountering vessels and length and speed of the oil tankers. Also the encounter type is determined, i.e. whether vessels overtake, cross or meet. The derived distributions are conditional to the direction of tanker travel and the respective GOFREP areas. The model includes the four steps outlined below.

Step 1: The AIS data is prepared for further analysis. In particular, the positions of all vessels are determined at examining times T_i, with an interval ΔT of 1 min, using linear interpolation from the AIS data points.
Step 2: The encounters between tankers with other vessels are detected. First, the trajectories of all tankers are extracted from the processed AIS data. Then, a circular inspection area is drawn around the tanker position at each examining time T_i, as in the method by van Dorp and Merrick (2011). While this inspection area is not the same concept as the ship domain[10], it is known that application of alternate ship domains leads to detection of different sets of encounters, particularly with respect to in the location where these occur (Goerlandt and Kujala, 2014). The uncertainty about the limits of the inspection area is here considered by applying alternative hypotheses (AH 1 in Fig. 4), setting the limit at 0.5 nm and 1 nm. Then, violations of the tanker inspection area with encountering vessels are detected by comparing AIS positions with the inspection area limits. The related information (vessel dimensions, types, speeds, courses) is stored.
Step 3: For each encounter, additional information is derived from the parameters in AIS data. The GOFREP area in which the encounter occurs is determined by comparing the positions of the tankers with the boundaries of the GOFREP sectors, see FTA (2010). The direction of tanker travel, i.e. whether the vessel is sailing into or out from the Gulf of Finland is derived from the course over ground of the tanker: courses toward east are inbound, courses toward west are outbound. Finally, the classification of the encounter type is based on a method proposed by Tam and Bucknall (2010), distinguishing overtaking, crossing and meeting encounters. The method classifies encounters based on the headings and bearings of both interacting vessels.
Step 4: The data is further processed and analyzed. Only one data field is retained per day, per GOFREP area, for an encounter between a unique pair of vessels (as identified by the MMSI number, see Section B.1.1). This is done to account for the fact that certain encounters last much longer than others, potentially skewing the results. Distributions for the length and speed of

[10] A ship domain is defined as "the surrounding effective waters which the navigators of a ship want to keep clear of other ships or fixed objects" (Qu et al., 2011), whereas an inspection domain is used only to determine which vessels get into each other's vicinity, without any reference to this area being indicative of a specific risk level (van Dorp and Merrick, 2011).

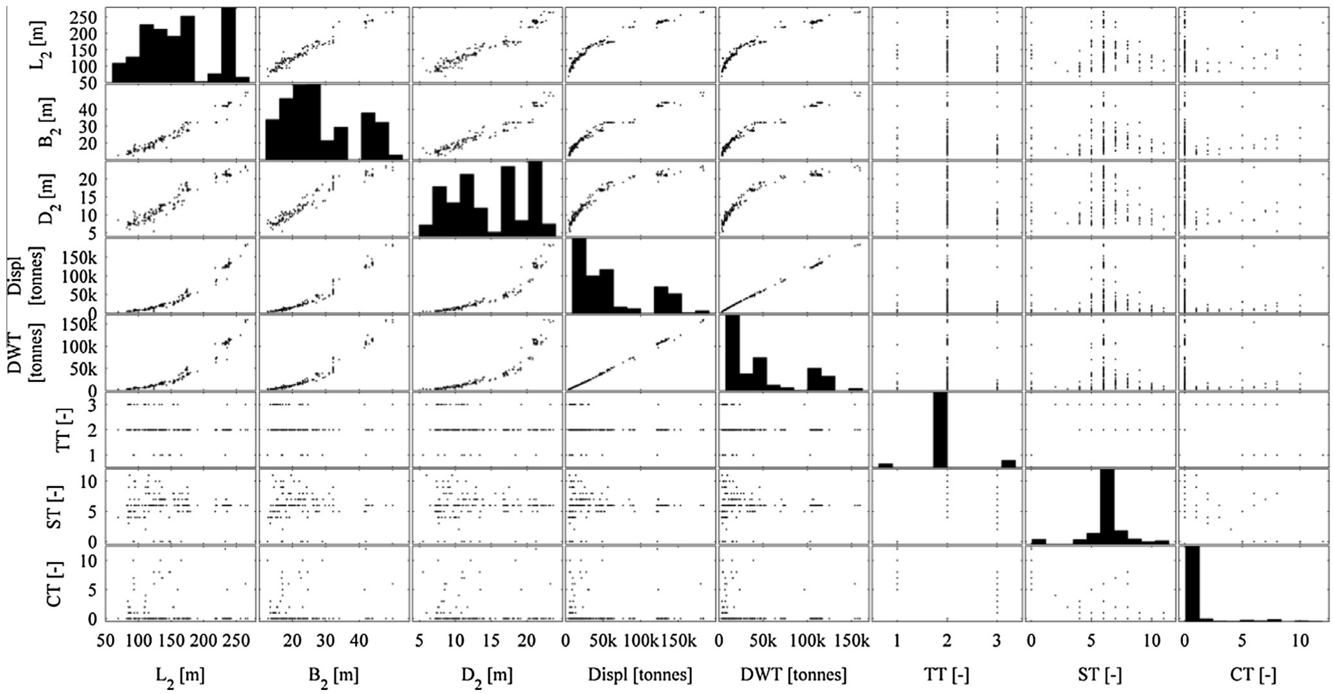

Fig. B1. Data for tankers considered in the analysis.

Table B4
Probabilities of tanker types in different sea areas, based on Nyman et al. (2010).

GoF area	*P* (oil product tanker)	*P* (crude oil tanker)
Finnish GOFREP	0.7	0.3
Estonian GOFREP	0.7	0.3
Russian national VTS	0.56	0.44

tankers and encountering vessels, as well as the type of encountering vessels are derived, conditional to the GOFREP area and the direction of travel. Finally, distributions of encounter type are derived from the obtained data set of vessel interactions.

B.3. Evidence for situational factors at collision impact

From collision accident reports and forensic analysis of ship collisions, it is known that the impact conditions typically differ from the encounter conditions (Cahill, 2002; Wang et al., 2013). Hence, the impact speeds differ from the encounter speeds, and the impact angle differs from the encounter angle. Furthermore, the impact conditions include additional situational factors such as which of the colliding vessels is the struck vessel and the damage location relative to the length of the tanker. However, as collisions are rare events and the collision evasive maneuvering is a complex phenomenon, there are important uncertainties related to the relationship between encounter and impact.

A number of models for establishing a relation between the relevant situational factors has been proposed, see Ståhlberg et al. (2013). In the BN model of Fig. 4, these models are considered as a set of alternative hypotheses. The models are outlined in Table B7, including a reference. V_1^I and V_2^I are the impact speeds of the encountering vessel and the tanker. V_1^E and V_2^E are the corresponding encounter speeds. The impact location l along the tanker hull is measured relative to the ship stern. φ^I and φ^E are respectively the impact and encounter angle, where φ^I is measured from the bow of the striking vessel. The table also contains assigned probabilities for alternative hypotheses AH 2–AH 6.

B.4. Evidence for probability of tanker collision

Various models have been proposed for estimating the probabilities of collisions in sea areas, e.g. Friis-Hansen and Simonsen (2002) and Weng et al. (2012). These methods follow the concept introduced by Fujii and Shiobara (1971), combining a count of encounters in an area with a causation probability, which is defined as the probability of failing to avoid a collision given a cri-

Table B5
Probabilities of tanker being laden in different sea areas, based on COWI (2011).

GoF area	Inbound		Outbound	
	Oil product	Crude oil	Oil product	Crude oil
Finnish GOFREP	0.5	DWT ⩽ 20 kton: 0.33 DWT ⩽ 80 kton: 0.084 DWT ⩽ 120 kton: 0.07 DWT > 120 kton: 0.06	DWT ⩽ 120 kton: 0.5 DWT > 120 kton: 0.99	DWT < 40 kton: 0.99 DWT < 120 kton: 0.98 DWT < 160 kton: 0.995
Estonian GOFREP	0.5	DWT ⩽ 20 kton: 0.33 DWT ⩽ 80 kton: 0.084 DWT ⩽ 120 kton: 0.07 DWT > 120 kton: 0.06	DWT ⩽ 120 kton: 0.5 DWT > 120 kton: 99	DWT < 40 kton: 0.99 DWT < 120 kton: 0.98 DWT < 160 kton: 0.995
Russian national VTS	DWT ⩽ 30 k: 0.5 DWT > 30 k: 0.01	0.01	DWT ⩽ 120 kton: 0.5 DWT > 120 kton: 0.99	0.99

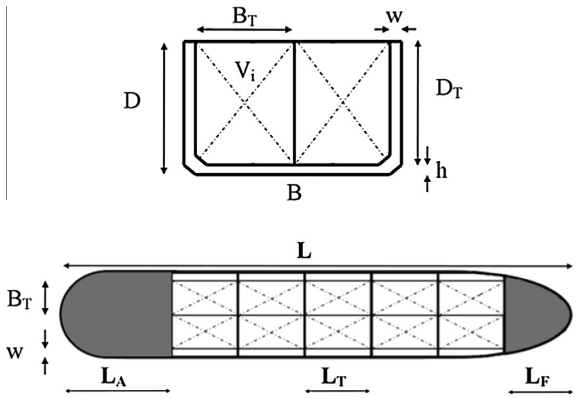

Fig. B2. Definition of tank dimensions and ship parameters.

Table B6
Probabilities of different bunker tank configurations, based on McAllister et al. (2003).

Vessel size (ktonnes)	Total bunker tank capacity (m^3)	Probability of bunker tank configuration					
		B1	B2	B3	B4	B5	B6
DWT < 10	500	1/2	1/2				
10 ⩽ DWT < 60	1000	1/2	1/2				
60 ⩽ DWT < 80	2000	1/3	1/3		1/3		
80 ⩽ DWT < 120	2500	1/3	1/3		1/3		
120 ⩽ DWT < 160	4000		1/4	1/4		1/4	1/4

tical ship encounter. However, the probability estimates have been found unreliable, due to questions related to the encounter detection mechanism (Goerlandt and Kujala, 2014) and the numerical values of the causation factors. Hence, there is uncertainty in applying these models (Sormunen et al., 2014).

As the focus of the case study concerns the consequence dimension, a crude evaluation of the collision occurrence is made based on available accident data and traffic information. Based on accident data analysis, an annualized collision frequency of 0.22 is found for the Gulf of Finland in open sea conditions, irrespective of ship type (Nyman et al., 2010). Furthermore, analysis of the number of vessels operating in the area indicates that between 13% (Montewka et al., 2010a) and 20% (Nyman et al., 2010) are oil product or crude oil tankers. It is known that maritime accidents are underreported to accident databases (Qu et al., 2012). Based on an analysis by Hassel et al. (2011), it is assumed that accidents in the Gulf of Finland are underreported at a 20% rate. Assuming that the accident probability is equal to all ship types, this leads to a probability of a collision involving a tanker between $3.58 \cdot 10^{-2}$ and $5.50 \cdot 10^{-2}$. These probabilities are taken as alternative hypothesis AH 7 in the BN. In terms of the rationale of Eq. (6), an equal probability is assigned to these.

B.5. Model for oil spill size

The model for the oil spill size integrates various data sources and engineering models: data concerning tanker cargo tank layout (Section B.1.3), a model for estimating the location and volumes of cargo tanks (Section B.1.4), data and information related to bunker tanks (Section B.1.5) and a model for estimating the damage size conditional to the impact conditions. This model is briefly introduced first. Then, the integrated procedure for determining the oil spill size is outlined.

B.5.1. Model for estimating the damage extent conditional to impact situation

A ship–ship collision is a complex, non-linear phenomenon which can be understood as a coupling of two dynamic processes. A first process concerns the redistribution of kinetic energy and its conversion to deformation energy due to two ship-shaped bodies coming into contact. This is coupled to a second process: the elastic and plastic deformation of the steel structures due to applied contact pressure. These processes are known as "outer dynamics" and "inner dynamics" (Terndrup Pedersen and Zhang, 1998). Several models have been proposed for evaluating hull damages in collisions, primarily aimed at ship design (Ehlers, 2011).

A relatively simple model, developed specifically for use in oil spill risk assessment, is presented in van de Wiel and van Dorp (2011). It is a statistical meta-model, based on a fit of a large set of damage extents as determined by a mechanical engineering model for a set of tanker designs (Brown and Chen, 2002). The model uses a set of dimensionless predictor variables x_i to estimate the damage length y_L and the penetration depth y_T, defined in Fig. B4:

$$y_L = exp\left(\sum_{i=1}^{5}\hat{\beta}_0^l + \sum_{j=1}^{5}\hat{\beta}_{i,j}^l x_j^i\right) \tag{B7}$$

$$y_T = exp\left(\sum_{i=1}^{5}\hat{\beta}_0^t + \sum_{j=1}^{5}\hat{\beta}_{i,j}^t x_j^i\right) \tag{B8}$$

The predictor variables x_i are functions of impact situational factors. x_1 and x_2 are derived from the perpendicular and tangential collision kinetic energy, which are functions of the vessel masses, impact speeds and impact angle. x_3 is a function of the relative damage location, x_4 of the bow entrance angle and x_5 is a function of the tanker length (for y_L) or the tanker width (for y_T). For expressions for the variables x_i and the regression coefficients $\hat{\beta}$, see van de Wiel and van Dorp (2011).

The model determines the maximum and minimum location of the longitudinal damage extent, respectively y_{L1} and y_{L2}, as a function of the damage length y_L, the tanker length L, the relative damage location l and the damage direction θ. The latter variable accounts for the phenomenon that the longitudinal damage extent will typically not be symmetrical around the impact location. The model for θ is conditional to the impact angle φ and the relative tangential velocity v_T.

B.5.2. Integrated model for oil spill size conditional to impact situation

In the BN of Fig. 4, the probabilities of oil spill sizes conditional to the impact situational variables are determined based on the following procedure:

i. The model of Section B.1.4 is used for all vessels to determine the cargo tank layout and volumes, and the mass in ballast condition, using the data of Section B.1.3.
ii. For $k = 1 \rightarrow 10$: Select a bunker tank layout, based on Section B.1.5.
iii. For all combinations of V_1^I, V_2^I, φ^I, l, η, M_{EV} and the tanker loading condition TLC:

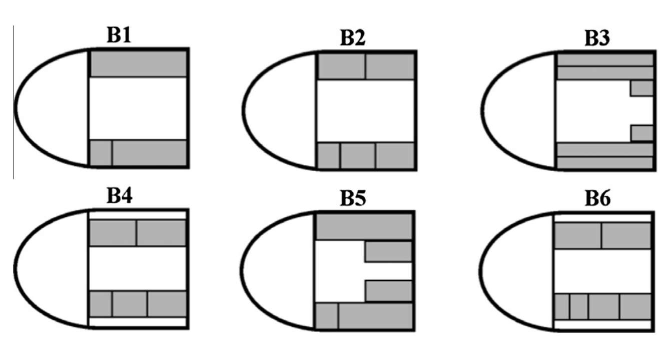

Fig. B3. Definition of bunker tank configurations, based on McAllister et al. (2003).

Table B7
Alternative hypotheses for impact situational factors.

Variable	Model	$p\ (M_i)$	Outline	Reference
AH 2: Encountering vessel impact speed V_1^I	M1	0.1	$V_1^I = V_1^E$	van Dorp and Merrick (2011)
	M2	0.1	$V_1^I = Tr\left(0, \frac{2}{3}V_1^E, V_1^E\right)$	COWI (2011)
	M3	0.1	$V_1^I = \begin{cases} U\left(0, 0.75V_1^E\right) \\ Tr\left(0.75V_1^E, V_1^E\right) \end{cases}$	Lützen (2001)
	M4	0.7	Empirical relation $V_1^I = f\left(V_1^E\right)$	Sormunen et al. (2014))
AH 3: Tanker impact speed V_2^I	M1	0.1	$V_2^I = V_2^E$	van Dorp and Merrick (2011)
	M2	0.1	$V_2^I = Tr\left(0, \frac{2}{3}V_2^E, V_2^E\right)$	COWI (2011)
	M3	0.1	$V_2^I = Tr\left(0, V_2^E\right)$	Lützen (2001)
	M4	0.7	Empirical relation $V_2^I = f(V_2^E)$	Sormunen et al. (2014)
AH 4: Impact location along tanker l	M1	0.25	$l = U(0, L_2)$	Rawson et al. (1998)
	M2	0.75	Empirical histogram	Samuelides et al. (2008)
AH 5: Impact angle φ^I	M1	0.05	$\varphi^I = \varphi^E$	van Dorp and Merrick (2011)
	M2	0.2	$\varphi^I = U(0, 180)$	Rawson et al. (1998)
	M3	0.1	$\varphi^I = Tr(0, \varphi^E, 180)$	Lützen (2001)
	M4	0.2	$\varphi^I = N(90, 29)$	NRC (2001)
	M5	0.05	CR and MT: $\varphi^I = U(30, 150)$ OT: $\begin{cases} p = 0.1 : \varphi^I = U(30, 150) \\ p = 0.9 : \varphi^I = \varphi^E \end{cases}$	COWI (2011)
	M6	0.4	Empirical histogram	Samuelides et al. (2008)
AH 6: Tanker striking or struck	M1	0.5	P (tanker struck) = 0.5	Klanac et al. (2010)
	M2	0.5	P (tanker struck struck) = 0.8	IMO (2008)

Notes: $U(a, b)$ = uniform distribution, $Tr(a, b, c)$ = triangular distribution, $N(\mu, \sigma)$ = normal distribution, other notations see Appendix A.

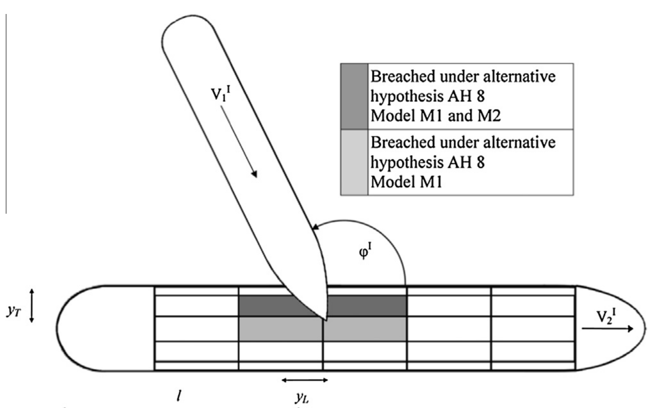

Fig. B4. Definition of impact situation, damage extent and hypotheses for breached compartments.

For $j = 1 \rightarrow 25$:

a. Sample a value in the considered discrete state of V_1^I, V_2^I, φ^I, l, η, M_{EV} and TLC.
b. Determine damage length y_L and damage depth y_T, see Section B.5.1.
c. Determine limits of the longitudinally breached area y_{L1} and y_{L2}, see Section B.5.1.
d. Compare y_{L1} and y_{L2} with the locations of the transverse bulkheads TBH, see Fig. B4.
e. Compare y_T with the locations of the longitudinal bulkheads LBH, see Fig. B4.
f. The volume of spilled oil is calculated as $V_{oil} = \sum_{i=1}^{N} V_i$, with N the number of tanks enclosed in the area encompassed by the comparisons in steps d and e, and V_i the cargo or bunker tank volume, see Sections B.1.4 and B.1.5.
g. The mass of spilled oil is calculated as $M_{oil} = \rho_{oil}V_{oil}$, assuming $\rho_{oil} = 0.9$ tonne/m^3.

Count the relative occurrence frequency of M_{oil} for each of the discrete classes of the BN-variable V_{25}, see Appendix A.

iv. The procedure ii.–iii. is repeated for all tanker designs.
v. The CPTs for the individual tanker designs are averaged over the group of vessels in the same deadweight range, according to the discretization in BN variable V_6, see Appendix A.
vi. Aggregation of all averaged CPTs for all tanker deadweights leads to the complete CPT for the BN variable V_{25}, see Appendix A.

Table B8
Alternative hypotheses for damage extent.

Variable	Model	$p\ (M_i)$	Outline
AH 8: Damage extent and oil spill size	M1	0.1	Damage extent is rectangular area covered by damage limits y_{L1}, y_{L2} and y_T All oil is spilled from the damaged compartments
	M2	0.9	Damage extent is rectangular area covered by damage limits y_{L1}, y_{L2} and y_T Maximum two longitudinal bulkheads breached Maximum two transverse bulkheads breached All oil is spilled from the damaged compartments
AH 9: Damage extent in relation to impact angle	M1	0.2	Hull breach occurs for all impact angles φ^I Damage extent and oil spill size according to AH 8
	M2	0.8	Hull breach occurs only for impact angles $\varphi^I \in [36°, 144]$ Damage extent and oil spill size according to AH 8

Table C1
Definition of test cases for evaluating influence of alternative hypotheses.

Test case	AH 1	AH 2	AH 3	AH 4	AH 5	AH 6	AH 7	AH 8	AH 9
C1	B	B	B	B	B	B	B	B	B
C2	M1	B	B	B	B	B	B	B	B
C3	M2	B	B	B	B	B	B	B	B
C4	B	M1	B	B	B	B	B	B	B
C5	B	M3	B	B	B	B	B	B	B
C6	B	B	M1	B	B	B	B	B	B
C7	B	B	M3	B	B	B	B	B	B
C8	B	B	B	M2	B	B	B	B	B
C9	B	B	B	M1	B	B	B	B	B
C10	B	B	B	B	M4	B	B	B	B
C11	B	B	B	B	M5	B	B	B	B
C12	B	B	B	B	B	M2	B	B	B
C13	B	B	B	B	B	M1	B	B	B
C14	B	B	B	B	B	B	M2	B	B
C15	B	B	B	B	B	B	M1	B	B
C16	B	B	B	B	B	B	B	M1	B
C17	B	B	B	B	B	B	B	M2	B
C18	B	B	B	B	B	B	B	B	M1
C19	B	B	B	B	B	B	B	B	M2
C20	M1	M1	M1	M2	M4	M2	M2	M1	M1
C21	M2	M3	M3	M1	M5	M1	M1	M2	M2

Notes: B: baseline weighing of alternative hypothesis, M1–M5: alternative models for various alternative hypotheses.

In this procedure, it is assumed that all oil in a tank is lost. More accurate models for evaluating the oil spill exist (Sergejeva et al., 2013; Tavakoli et al., 2010), but their application requires information regarding the damage height and opening size and shape. As the damage extent model by van de Wiel and van Dorp (2011) only provides crude information regarding the limits of the breached area, no detailed information for these factors is available. Thus, the conservative assumption of a complete loss of oil from damaged tanks is made, as in van Dorp and Merrick (2011).

B.5.3. Alternate hypotheses

In the procedure outlined in Section B.5.2, several assumptions are made, leading to epistemic uncertainties. These concern the damage extent in relation to the bow penetration depth and the damage extent in relation to the impact angle. These epistemic uncertainties are considered through a set of alternative hypotheses.

In steps iii./d.–e. of the procedure in Section B.5.2, the damage extent is taken as the rectangular area covered by the damage limits y_{L1}, y_{L2} and y_T, in line with the application by van de Wiel and van Dorp (2011). It is however known that the breach of each additional bulkhead corresponds to an increased amount of required deformation energy (Ehlers and Tabri, 2012; Ståhlberg et al., 2013). As the model by Brown and Chen (2002), on which the meta-model by van de Wiel and van Dorp (2011) is based, only accounts for the outer hull and the double hull bulkhead, there is uncertainty regarding the breach of the subsequent longitudinal bulkhead(s). Likewise, the breach of transverse bulkheads is uncertain, as the bulkhead structure requires more deformation energy for a breach to occur (Klanac et al., 2010). This uncertainty is considered through model M2 for AH 8, as outlined in Table B8. The same procedure as in Section B.5.2 is followed for constructing the CPT for BN-variable V_{26} (see Appendix A), but maximum two longitudinal and two transverse bulkheads can be breached.

Another uncertainty concerns the collision consequences under oblique impact angles. While advanced theoretical models exist for non-perpendicular collisions (Tabri et al., 2010), the behavior of the interface between the impacted structures under oblique angles is not well understood in the present state of research. Some models assume that the ship hulls will only grind alongside, without leading to extensive structural damage (COWI, 2011; Ståhlberg et al., 2013). In such cases, no breach of the double hull structure is assumed, and no oil is spilled. In other analyses, it is taken that all impact angles lead to hull breach (van Dorp and Merrick, 2011). In the BN, this uncertainty is considered through AH 9, as outlined in Table B8.

Appendix C. Test cases for alternative hypotheses

In Table C1, the test cases for quantifying the influence of the alternative hypotheses AH 1–AH 9 in the BN of Fig. 4 are summarized. In each test case, the AHs are either taken at the baseline level or set at a particular model alternative. The model is run for each of these cases, and summary statistics of the corresponding results are shown in Fig. 6.

References

Akhtar, J., Bjornskau, T., Jean-Hansen, V., 2012. Oil spill risk analysis of routeing heavy ship traffic in Norwegian waters. WMU J. Marit. Aff. 11, 233–247. http://dx.doi.org/10.1007/s13437-011-0016-8.

Apostolakis, G.E., 2004. How useful is quantitative risk assessment? Risk Anal. 24, 515–520.

Aven, T., 2007. On the ethical justification for the use of risk acceptance criteria. Risk Anal. 27, 303–312.

Aven, T., 2008. Risk Analysis: Assessing Uncertainties beyond Expected Values and Probabilities. Wiley, Chichester, UK.

Aven, T., 2009. Perspectives on risk in a decision-making context – review and discussion. Saf. Sci. 47, 798–806. http://dx.doi.org/10.1016/j.ssci.2008.10.008.

Aven, T., 2010a. On how to define, understand and describe risk. Reliab. Eng. Syst. Saf. 95, 623–631. http://dx.doi.org/10.1016/j.ress.2010.01.011.

Aven, T., 2010b. Misconceptions of Risk. John Wiley & Sons Ltd., Chichester, UK.

Aven, T., 2010c. Some reflections on uncertainty analysis and management. Reliab. Eng. Syst. Saf. 95, 195–201. http://dx.doi.org/10.1016/j.ress.2009.09.010.

Aven, T., 2011a. Interpretations of alternative uncertainty representations in a reliability and risk analysis context. Reliab. Eng. Syst. Saf. 96, 353–360.

Aven, T., 2011b. Selective critique of risk assessments with recommendations for improving methodology and practise. Reliab. Eng. Syst. Saf. 96, 509–514. http://dx.doi.org/10.1016/j.ress.2010.12.021.

Aven, T., 2012. The risk concept—historical and recent development trends. Reliab. Eng. Syst. Saf. 99, 33–44. http://dx.doi.org/10.1016/j.ress.2011.11.006.

Aven, T., 2013. Practical implications of the new risk perspectives. Reliab. Eng. Syst. Saf. 115, 136–145. http://dx.doi.org/10.1016/j.ress.2013.02.020.

Aven, T., Guikema, S., 2011. Whose uncertainty assessments (probability distributions) does a risk assessment report: the analysts' or the experts'? Reliab. Eng. Syst. Saf. 96, 1257–1262. http://dx.doi.org/10.1016/j.ress.2011.05.001.

Aven, T., Heide, B., 2009. Reliability and validity of risk analysis. Reliab. Eng. Syst. Saf. 94, 1862–1868. http://dx.doi.org/10.1016/j.ress.2009.06.003.

Aven, T., Reniers, G., 2013. How to define and interpret a probability in a risk and safety setting. Saf. Sci. 51, 223–231. http://dx.doi.org/10.1016/j.ssci.2012.06.005.

Aven, T., Renn, O., 2009. On risk defined as an event where the outcome is uncertain. J. Risk Res. 12, 1–11.

Aven, T., Vinnem, J.E., 2005. On the use of risk acceptance criteria in the offshore oil and gas industry. Reliab. Eng. Syst. Saf. 90, 15–24. http://dx.doi.org/10.1016/j.ress.2004.10.009.

Aven, T., Zio, E., 2011. Some considerations on the treatment of uncertainties in risk assessment for practical decision making. Reliab. Eng. Syst. Saf. 96, 64–74. http://dx.doi.org/10.1016/j.ress.2010.06.001.

Aven, T., Zio, E., 2014. Foundational issues in risk assessment and risk management. Risk Anal. 34, 1164–1172.

Ayyub, B.M., 2001. Elicitation of Expert Opinions for Uncertainty and Risk. CRC Press, Boca Raton, FL.

Bi, H., Si, H., 2012. Dynamic risk assessment of oil spill scenario for Three Gorges Reservoir in China based on numerical simulation. Saf. Sci. 50, 1112–1118.

Brown, A.J., 2002. Collision scenarios and probabilistic collision damage. Mar. Struct. 15, 335–364. http://dx.doi.org/10.1016/S0951-8339(02)00007-2.

Brown, A.J., Chen, D., 2002. Probabilistic method for predicting ship collision damage. Ocean Eng. Int. 6, 54–65.

Brunila, O.-P., Storgård, J., 2012. Oil Transportation in the Gulf of Finland in 2020 and 2030 (No. A61). Publications from the Centre for Maritime Studies. University of Turku, Turku.

Cahill, R.A., 2002. Collisions and their Causes, third ed. The Nautical Institute, London.

Carmines, E.G., Zeller, R.A., 1979. Reliability and Validity Assessment, Quantitative Applications in the Social Sciences. SAGE Publications, Thousand Oaks, CA.

Chen, D., 2000. Simplified Ship Collision Model (M.Sc.). Virginia Polytechnic Institute and State University, Blacksburg, Virginia.
Coupé, V.M.H., van der Gaag, L.C., 2002. Properties of sensitivity analysis of Bayesian Belief networks. Ann. Math. Artif. Intell. 36, 323–356.
COWI, 2011. BRISK – Sub-regional Risk of Spill of Oil and Hazardous Substances in the Baltic Sea.
Dalton, T., Jin, D., 2010. Extent and frequency of vessel oil spills in US marine protected areas. Mar. Pollut. Bull. 60, 1939–1945.
De Rocquigny, E., Devictor, N., Tarantola, S. (Eds.), 2008. Uncertainty in Industrial Practice – A Guide to Quantitative Uncertainty Management. Wiley, New Jersey.
Debnath, A.K., Chin, H.C., 2010. Navigational traffic conflict technique: a proactive approach to quantitative measurement of collision risks in port waters. J. Navigation 63, 137–152.
Diekmann, S., Peterson, M., 2013. The role of non-epistemic values in engineering models. Sci. Eng. Ethics 19, 207–218.
Douglas, H., 2009. Science, Policy and the Value-free Ideal. University of Pittsburgh Press, Pittsburgh, PA.
Ehlers, S., 2011. A review of collision and grounding damage assessment method. Mar. Syst. Ocean Technol. 6, 5–15.
Ehlers, S., Tabri, K., 2012. A combined numerical and semi-analytical collision damage assessment procedure. Mar. Struct. 28, 101–119. http://dx.doi.org/10.1016/j.marstruc.2012.05.005.
Evangelista, J., 2002. Scaling the tanker market. Surv. – Q. Mag. ABS, 5–11.
Faber, M.H., 2005. On the treatment of uncertainties and probabilities in engineering decision analysis. J. Offshore Mech. Arctic Eng. 127, 243–248.
Felski, A., Jaskolski, K., 2013. The integrity of information received by means of AIS during anti-collision maneuvering. TransNav – Int. J. Mar. Navig. Saf. Sea Transp. 7, 95–100.
Fenton, N., Neil, M., 2012. Risk Assessment and Decision Analysis with Bayesian Networks. CRC Press.
Flage, R., Aven, T., 2009. Expressing and communicating uncertainty in relation to quantitative risk analysis (QRA). Reliab. Risk Anal. Theory Appl. 2, 9–18.
Forrester, J.W., Senge, P.M., 1980. Tests for building confidence in system dynamics models. TIMS Stud. Manage. Sci. 14, 209–228.
Friis-Hansen, P., Simonsen, B.C., 2002. GRACAT: software for grounding and collision risk analysis. Mar. Struct. 15, 383–401.
FTA, 2010. GOFREP Master's Guide.
Fujii, Y., Shiobara, R., 1971. The analysis of traffic accidents – studies in marine traffic accidents. J. Navigation 24, 534–543.
Garcia Negro, M.C., Villasante, S., Carballo Penela, A., Rodriguez Rodriguez, R., 2009. Estimating the economic impact of the Prestige oil spill on the Death Coast (NW Spain) fisheries. Mar. Policy 33, 8–23.
Goerlandt, F., Kujala, P., 2014. On the reliability and validity of ship–ship collision risk analysis in light of different perspectives on risk. Saf. Sci. 62, 348–365. http://dx.doi.org/10.1016/j.ssci.2013.09.010.
Goerlandt, F., Montewka, J., 2015. Maritime transportation risk analysis: review and analysis in light of some foundational issues. Reliab. Eng. Syst. Safe. 138, 115–134. http://dx.doi.org/10.1016/j.ress.2015.01.025.
Graveson, A., 2004. AIS – an inexact science. J. Navigation 57, 339–343.
Gucma, L., Przywarty, M., 2008. The model of oil spills due to ship collisions in Southern Baltic Area. TransNav Int. J. Mar. Navig. Saf. Sea Transp. 2, 415–419.
Hampel, J., 2006. Different concepts of risk – a challenge for risk communication. Int. J. Med. Microbiol. 296, 5–10.
Hänninen, M., 2014. Bayesian networks for maritime traffic accident prevention: benefits and challenges. Accid. Anal. Prev. 73, 305–312.
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 Reliab. 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, 235–247. http://dx.doi.org/10.1016/S0925-7535(98)00048-4.
Hassel, M., Asbjørnslett, B.E., Hole, L.P., 2011. Underreporting of maritime accidents to vessel accident databases. Accid. Anal. Prev. 43, 2053–2063.
Hassler, B., 2011. Accidental versus operational spills from shipping in the Baltic Sea: risk governance and management strategies. Ambio 40, 170–178.
HELCOM, 2014. HELCOM – Baltic Sea Trends – Accidents. <http://helcom.fi/baltic-sea-trends/maritime/accidents> (accessed 12.20.14).
Helle, I., Lecklin, T., Jolma, A., Kuikka, S., 2011. Modeling the effectiveness of oil combating from an ecological perspective – a Bayesian network for the Gulf of Finland; the Baltic Sea. J. Hazard. Mater. 185, 182–192.
Hermansson, H., 2012. Defending the conception of "objective" risk. Risk Anal., 32
Hodges, J.S., 1991. Six (or so) things you can do with a bad model. Oper. res. 39, 355–365.
Hogström, P., 2012. RoPax Ship Collision – A Methodology for Survivability Analysis (PhD Dissertation). Chalmers University of Technology, Gothenburg, Sweden.
IHS Maritime, 2013. Maritime Insight and Information.
IMO, 2003. Revised Interim Guidelines for the Approval of Alternative Methods of Design and Construction of Oil Tankers.
IMO, 2008. Formal Safety Assessment – FSA Crude Oil Tankers.
IMO, 2010. Manual on Oil Spill Risk Evaluation and Assessment of Response Preparedness. IMO Publishing.
Jolma, K., Haapasaari, H., 2014. Oil and Chemical Spill Response in Finland – Current Situation.
Kaplan, S., 1997. The words of risk analysis. Risk Anal. 17, 407–417. http://dx.doi.org/10.1111/j.1539-6924.1997.tb00881.x.
Klanac, A., Varsta, P., 2011. Design of marine structures with improved safety for environment. Reliab. Eng. Syst. Saf. 96, 75–90. http://dx.doi.org/10.1016/j.ress.2010.06.016.
Klanac, A., Duletić, T., Erceg, S., Ehlers, S., Goerlandt, F., Frank, D., 2010. Environmental risk of collisions in the enclosed European waters: Gulf of Finland, Northern Adriatic and the implications for tanker design. In: 5th International Conference on Collision and Grounding of Ships. Aalto University, Espoo, Finland, pp. 55–65.
Kloprogge, P., van der Sluijs, J.P., Petersen, A.C., 2011. A method for the analysis of assumptions in model-based environmental assessments. Environ. Model. Softw. 26, 289–301.
Koller, D., Friedman, N., 2009. Probabilistic graphical models: principles and techniques,, .. Adaptive Computation and Machine Learning, first ed. The MIT Press.
Lecklin, T., Ryöma, R., Kuikka, S., 2011. A Bayesian network for analyzing biological acute and long-term impacts of an oil spill in the Gulf of Finland. Mar. Pollut. Bull. 62, 2822–2835.
Lee, M., Jung, J.-Y., 2013. Risk assessment and national measure plan for oil and HNS spill accidents near Korea. Mar. Pollut. Bull. 73, 339–344.
Lehikoinen, A., Luoma, E., Mäntyniemi, S., Kuikka, S., 2013. Optimizing the recovery efficiency of Finnish oil combating vessels in the Gulf of Finland using Bayesian Networks. Environ. Sci. Technol. 47, 1792–1799. http://dx.doi.org/10.1021/es303634f.
Li, S., Meng, Q., Qu, X., 2012. An overview of maritime waterway quantitative risk assessment models. Risk Anal. 32, 496–512.
Lindley, D.V., 2006. Understanding Uncertainty, 11th ed. Wiley-Interscience, Hoboken, NJ.
Lützen, M., 2001. Ship Collision Damage. PhD Thesis, Department of Mechanical Engineering, Technical University of Denmark, Lyngby.
McAllister, T., Rekart, C., Michel, K., 2003. Evaluation of Accidental Oil Spills from Bunker Tanks (Phase I) (Technical Report No. SSC-424). Ship Structures Committee, Washington, DC.
Miraglia, R.A., 2002. The cultural and behavioral impact of the Exxon Valdez oil spill on the native peoples of the Prince William Sound, Alaska. Spill Sci. Technol. Bull. 7, 75–87.
Montewka, J., Hinz, T., Kujala, P., Matusiak, J., 2010a. Probability modelling of vessel collisions. Reliab. Eng. Syst. Saf. 95, 573–589. http://dx.doi.org/10.1016/j.ress.2010.01.009.
Montewka, J., Ståhlberg, K., Seppala, T., Kujala, P., 2010. Elements of risk analysis for collision of oil tankers. In: Reliability, Risk and Safety: Back to the Future. Presented at the European Safety and Reliability Annual Conference, ESREL2010, Rhodes, Greece, pp. 1005–1013.
Montewka, J., Weckström, M., Kujala, P., 2013. A probabilistic model estimating oil spill clean-up costs – a case study for the Gulf of Finland. Mar. Pollut. Bull. 76, 61–71. http://dx.doi.org/10.1016/j.marpolbul.2013.09.031.
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. Reliab. Eng. Syst. Saf. 124, 142–157. http://dx.doi.org/10.1016/j.ress.2013.11.014.
Montewka, J., Goerlandt, F., Kujala, P., 2014b. On a systematic perspective on risk for formal safety assessment (FSA). Reliab. Eng. Syst. Saf. 127, 77–85. http://dx.doi.org/10.1016/j.ress.2014.03.009.
Mou, J.M., van der Tak, C., Ligteringen, H., 2010. Study on collision avoidance in busy waterways by using AIS data. Ocean Eng. 37, 483–490.
NRC, 2001. Environmental Performance of Tanker Designs in Collision and Grounding. Special Report 259. National Research Council.
Nyman, T., Porthin, M., Karppinen, S., 2010. Collision and Grounding Frequency Analyses in the Gulf of Finland.
OCIMF, 2008. Tanker Management Self-Assessment (TMSA). Oil Companies International Maritime Forum, London.
O'Hagan, A., Buck, C.E., Daneshkhak, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T., 2006. Uncertain Judgments: Eliciting Experts' Probabilities, first ed. Wiley.
Okumoto, Y., Takeda, Y., Mano, M., Okada, T., 2009. Design of Ship Hull Structures: A Practical Guide for Engineers, first ed. Springer.
Papanikolaou, A. (Ed.), 2009. Risk-based Ship Design: Methods, Tools and Applications. Springer.
Pitchforth, J., Mengersen, K., 2013. A proposed validation framework for expert elicited Bayesian Networks. Expert Syst. Appl. 40, 162–167.
Porathe, T., Lützhöft, M., Praetorius, G., 2013. Communicating intended routes in ECDIS: evaluating technological change. Accid. Anal. Prev. 60, 366–370.
Psaraftis, H.N., 2012. Formal safety assessment: an updated review. J. Mar. Sci. Technol. 17, 390–402.
Qu, X., Meng, Q., 2012. The economic importance of the Straits of Malacca and Singapore: an extreme-scenario analysis. Transport. Res. E: Logist. Transport. Rev. 48, 258–265.
Qu, X., Meng, Q., Suyi, L., 2011. Ship collision risk assessment for the Singapore Strait. Accid. Anal. Prev. 43, 2030–2036. http://dx.doi.org/10.1016/j.aap.2011.05.022.
Qu, X., Meng, Q., Li, S., 2012. Analyses and implications of accidents in Singapore Strait. Transport. Res. Rec. J. Transp. Res. Board 2273, 106–111.
Rawson, C., Crake, K., Brown, A., 1998. Assessing the environmental performance of tankers in accidental grounding and collision. SNAME Trans. 106, 41–58.

Rosqvist, T., 2010. On the validation of risk analysis – a commentary. Reliab. Eng. Syst. Saf. 95, 1261–1265.
Rosqvist, T., Tuominen, R., 2004. Qualification of formal safety assessment: an exploratory study. Saf. Sci. 42, 99–120.
Samuelides, M.S., Tabri, K., Incecik, A., Dimou, D., 2008. Scenarios for the assessment of the collision behavior of ships. Int. Shipbuild. Prog. 55, 145–162. http://dx.doi.org/10.3233/ISP-2008-0043.
Sergejeva, M., Laarnearu, J., Tabri, K., 2013. Hydraulic modelling of submerged oil spill including tanker hydrostatic overpressure. In: Analysis and Design of Marine Structures. CRC Press, Taylor and Francis Group.
Shrader-Frechette, K.S., 1991. Risk and Rationality: Philosophical Foundations for Populist Reforms. University of California Press, Berkeley.
Smailys, V., Česnauskis, M., 2006. Estimation of expected cargo oil outflow from tanker involved in casualty. Transport 21, 293–300.
Solberg, O., Njå, O., 2012. Reflections on the ontological status of risk. J. Risk Res. 15, 1201–1215.
Sormunen, O.-V.E., Goerlandt, F., Häkkinen, J., Posti, A., Hänninen, M., Montewka, J., Ståhlberg, K., Kujala, P., 2014. Uncertainty in maritime risk analysis: extended case study on chemical tanker collisions. Proc. Inst. Mech. Eng. Part M: J. Eng. Marit. Environ. http://dx.doi.org/10.1177/1475090213515640.
Ståhlberg, K., 2010. Estimating Deformation Energy in Ship–Ship Collisions with Stochastic Modeling (MSc thesis). Aalto University, Otaniemi, Finland.
Ståhlberg, K., Goerlandt, F., Ehlers, S., Kujala, P., 2013. Impact scenario models for probabilistic risk-based design for ship–ship collision. Mar. Struct. 33, 238–264. http://dx.doi.org/10.1016/j.marstruc.2013.06.006.
Tabri, K., 2010. Influence of coupling in the prediction of ship collision damage. In: Presented at the International Conference on Collision and Grounding of Ships, Aalto University, Espoo, Finland, pp. 133–138.
Tabri, K., Varsta, P., Matusiak, J., 2010. Numerical and experimental motion simulations of non-symmetric ship collisions. J. Mar. Sci. Technol. 15, 87–101.
Tam, C., Bucknall, R., 2010. Collision risk assessment for ships. J. Mar. Sci. Technol. 15, 257–270.
Tavakoli, M.T., Amdahl, J., Leira, B., 2010. Analytical and numerical modelling of oil spill from a side damaged tank. In: Proceedings of the Fifth International Conference on Collision and Grounding, ICCGS, Aalto University, Espoo, Finland.
Taylor, E., Steen, A., Meza, M., Couzigou, B., Hodges, M., Maranda, D., Ramos, J., Moyano, M., 2008. IOSC workshop report: a proposed international guide for oil spill response planning and readiness assessment. In: Presented at the International Oil Spill Conference, Savannah, Georgia, pp. 1–18.
Terndrup Pedersen, P., Zhang, S., 1998. On impact mechanics in ship collisions. Mar. Struct. 11, 429–449.
Thompson, P.B., Dean, W.R., 1996. Competing conceptions of risk. Risk Health Saf. Environ. 7, 361–384.
Trochim, W., Donnely, J.P., 2008. The Research Methods Knowledge Base, third ed. Atomic Dog Publishing.
USCG, 2012. Automatic Identification System – Encoding Guide.
Uusitalo, L., 2007. Advantages and challenges of Bayesian networks in environmental modelling. Ecol. Model. 203, 312–318.
Van de Wiel, G., van Dorp, J.R., 2011. An oil outflow model for tanker collisions and groundings. Ann. Oper. Res. 187, 279–304. http://dx.doi.org/10.1007/s10479-009-0674-5.
Van Dorp, J.R., Merrick, J.R., 2011. On a risk management analysis of oil spill risk using maritime transportation system simulation. Ann. Oper. Res. 249–277. http://dx.doi.org/10.1007/s10479-009-0678-1.
Vanem, E., Endresen, O., Skjong, R., 2008. Cost-effectiveness criteria for marine oil spill preventive measures. Reliab. Eng. Syst. Saf. 93, 1354–1368.
Vareman, N., Persson, J., 2010. Why separate risk assessors and risk managers? Further external values affecting the risk assessor qua risk assessor. J. Risk Res. 13, 687–700.
Wandall, B., 2004. Values in science and risk assessment. Toxicol. Lett. 152, 265–272.
Wang, Y., Zhang, J., Chen, X., Chu, X., Yan, X., 2013. A spatial–temporal forensic analysis for inland-water ship collisions using AIS data. Saf. Sci. 57, 187–202.
Watson, S.R., 1994. On the meaning of probability in probabilistic safety analysis. Reliab. Eng. Syst. Saf. 45, 261–269.
Weng, J., Meng, Q., Qu, X., 2012. Vessel collision frequency estimation in the Singapore Strait. J. Navigation 65, 207–221.
Wiśniewski, K., Kołakowski, P., 2003. The effect of selected parameters on ship collision results by dynamic FE simulations. Finite Elem. Anal. Des. 39, 985–1006.
Yan, X., Zhang, J., Zhang, D., Soares, Carlos Guedes, 2014. Challenges and developments in navigational risk assessment with large uncertainty. In: ASME 2014 33rd International Conference on Ocean, Offshore and Arctic Engineering, San Francisco, CA, p. 8. http://dx.doi.org/10.1115/OMAE2014-23411.
Zio, E., Apostolakis, G.E., 1996. Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive waste repositories. Reliab. Eng. Syst. Saf. 54, 225–241.