



Enhancing human performance in ship operations by modifying global design factors at the design stage



Jakub Montewka^{a,b,c,*}, Floris Goerlandt^a, Gemma Innes-Jones^d, Douglas Owen^d, Yasmine Hifi^e, Romanas Puisa^e

^a Aalto University, Department of Applied Mechanics, Marine Technology, Research Group on Maritime Risk and Safety, 02150 Espoo, Finland

^b Finnish Geospatial Research Institute, 02430 Masala, Finland

^c Gdynia Maritime University, Faculty of Navigation, Department of Transport and Logistics, 81-225 Gdynia, Poland

^d Lloyd's Register, EMEA, Bristol, UK

^e Brookes Bell R & D, Glasgow, UK

ARTICLE INFO

Keywords:

Human performance
Ship design
Collision
Grounding
Bayesian Belief Networks

ABSTRACT

Usually the improvements of human performance in the course of ship design process is carried out by modifying local ergonomics, like electronic visualisation and information display systems on the bridge or in the engine control room, stair or hatch covers design. However, the effect of global design factors (GDFs), such as ship motion, whole body vibration and noise, on human performance has not been given attention before. Such knowledge would allow the improvements of human performance by effective design modification on very early stage of ship design process.

Therefore, in this paper we introduce probabilistic models linking the effect of GDFs with the human performance suitable for ship design process. As a theoretical basis for modelling human performance the concept of Attention Management is utilized, which combines the theories described by Dynamic Adaptability Model, Cognitive Control Model and Malleable Attentional Resources Theory.

Since the analysed field is characterised by a high degree of uncertainty, we adopt a specific modelling technique along with a validation framework that allows uncertainty treatment and helps the potential end-users to gain confidence in the models and the results that they yield. The proposed models are developed with the use Bayesian Belief Networks, which allows systematic translation of the available background knowledge into a coherent network and the uncertainty assessment and treatment.

The obtained results are promising as the models are responsive to changes in the GDF nodes as expected. The models may be used as intended by naval architects and vessel designers, to facilitate risk-based ship design.

1. Introduction

Reduced human performance is reported as one of major factors contributing to the maritime accidents, [1–4]. In the recent years the studies related to the quantification of human performance for various shipborne operations have been gaining an increasing attention, resulting in a number of models and approaches, see for example [5–8]. At the same time, significant efforts have been made to study and implement local design modifications of ships improving the ergonomics thus human performance on board a ship and ultimately ship safety, [9–13].

A major recent advance in the field of maritime safety is the development of risk-based ship design methodology (RBSD), resulting in development of larger and potentially safer ships, [14,15]. Within

RBSD the assessment of the risk level with respect to predefined types of accidents is conducted in the early design stage, where a design modification is easy and cost-effective and risk is treated as a design objective, [12]. In risk analysis two aspects of the analysed accident are covered, its likelihood (accident prevention) and the anticipated consequences (accident mitigation). The latter is addressed by improving technical and structural reliability of a ship and it has been extensively studied over past years, [16–20]. The former is addressed usually by improving performance of a human, however the research on the effect of the overall ship design on human performance is in its infancy, [21–24], despite its relevance to the field of ship design, [25]. Such a method could be incorporated into the RBSD, improving human performance through modification of appropriate global design factors (GDFs), thus reducing the risk of accidents already at the early design

* Corresponding author at: Aalto University, Department of Applied Mechanics, Marine Technology, Research Group on Maritime Risk and Safety, 02150 Espoo, Finland.

<http://dx.doi.org/10.1016/j.ress.2016.11.009>

Received 9 February 2016; Received in revised form 26 August 2016; Accepted 17 November 2016

Available online 23 November 2016

0951-8320/© 2016 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Nomenclature		GTT	Generic Task Type
AMC	Attention Management Capability	HEP	Human Error Probability
AH	Alternative Hypothesis	HRA	Human Reliability Assessment
BBN	Bayesian Belief Networks	MART	Malleable Attentional Resources Theory
CCM	Cognitive Control Model	MII	Motion Induced Interruptions
DAAct	a triplet of Detection, Assessment, Action	MIS	Motion Induced Sickness
APOA	Assessed Proportion of Affect	NARA	Nuclear Action Reliability Assessment
DAM	Dynamic Adaptability Model	OOW	Officer Of the Watch
EPC	Error Producing Conditions	RBSD	Risk-Based Ship Design
GDFs	Global Design Factors (noise, whole body vibration, motion)	SCT	Safety Critical Tasks
		WBV	Whole Body Vibration

stage.

However, the possible risk reduction remains unknown until now, since the common human error quantification frameworks do not readily account for the specific effect of the GDFs.

Therefore, this paper presents advances, focusing on modelling the effect of GDFs on human performance which is measured with the probability of ship-ship collision and ship grounding, which in turn is a recognized proxy for risk in the RBSD process and can easily be incorporated therein.

As a results of extensive literature survey on the effects of human exposure to the following three GDFs: ship motion, noise and vibration see [24–26] a workable approach has emerged for modelling human performance focussing on attention management, which is found suitable for the given purpose. It is based on three theories: the Dynamic Adaptability Model, [29] Cognitive Control Model [30] and Malleable Attentional Resources Theory [31].

These foundations are used as a guide for constructing two models presented here, which are developed using Bayesian Belief Networks (BBNs). BBNs is capable of representing background knowledge about the analysed accidents, the evaluation of associated uncertainties, efficient reasoning and updating in light of new evidence, see for example [6,32–37].

Finally the models are validated adopting a framework as proposed by [38], which is found suitable for a given purpose, [39]. The framework allows for rigorous checks of the models along with the evaluation of uncertainty. As result, the models are found to behave in

response to GDF inputs as intended, and the obtained results are found valid for a given purpose. The models can offer a valid comparative assessment of ship designs with respect to human performance, which is their primary intention.

The remainder of the paper is organized as follows: Section 2 introduces the modelling framework, upon which the human performance models are developed. The models development process is presented in Section 3. Section 4 discusses the validation of the models and Section 5 concludes the paper.

2. Structure of the models

The aim of the presented models is two-fold. First they quantify the human performance in the presence of motion, noise and whole body vibration that are specific for each ship designs. Second they allow differentiation of various ship designs with respect to the human performance.

In this section we introduce a modelling framework adopted here that fits requirements of RBSD and recommendations by IMO regarding human performance modelling in maritime [40]. Subsequently the causal pathway linking the GDFs with human performance is described.

2.1. Modelling framework

To develop the models we adopt a generic modelling framework by

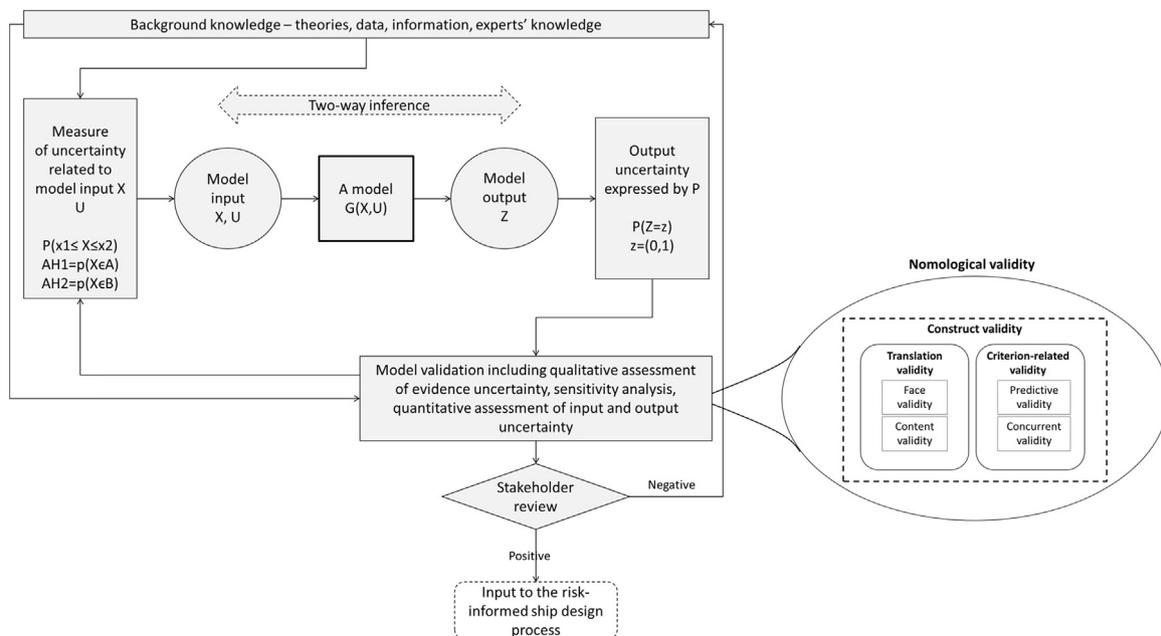


Fig. 1. A modelling framework adopted for human performance evaluation subjected to GDFs, suitable for risk-informed ship design.

[41], with modification, as depicted in Fig. 1. The framework aims to:

- integrate in systematic way the available background knowledge on the subject of the analysis,
- provide sound basis for the procedure of human performance assessment subjected to GDFs
- account for the existing uncertainty by determining its bound for the models,
- validate the developed probabilistic models,
- facilitate the decision making process.

The relevant data, theories, models and expert's judgment is obtained and gathered into a workable model, mimicking the described phenomenon. Since the level of available background knowledge significantly varies across the models, the modelling choice has to reflect on that, being able to:

- account for the uncertainty,
- propagate it through the model,
- infer in the presence of uncertainty,

Considering the above, we use probabilistic casual models, aka. Bayesian Belief Networks, as a modelling technique. BBNs is recognized tool for reasoning under uncertainty, linking various types of data into one whole for the wide uncertainty assessment and treatment. Within the BBNs the uncertainty is measured through the probabilities, and propagated through the model with the use of Bayesian Theory. The inherent feature of BBNs, two-ways reasoning, allows not only forward propagation of the evidences resulting in an outcome, but it is also possible to propagate backwards the evidences to estimate the most probable input variables, given a selected state of the output. The high-level modelling framework adopted here is depicted in Fig. 1 and elaborated in the following sections.

2.2. Linking the effects of GDFs with human performance

To describe the process through which exposure to GDFs causally affects the performance of a crew member in relation to specific operations (collision avoidance, grounding avoidance, tasks related to maintenance of ship technical systems) a causal pathway was developed through the mediating agent of the crewmember. The process serves to do three things:

1. Represent the mechanism by which GDFs exposure impacts human performance in operations.
2. Describe the overall topography of the final model.
3. Facilitate the identification of nodes.

GDFs can be considered a type of performance shaping factors (PSF), where PSFs are an aspect of the human's individual characteristics, environment, organisation, or task that specifically decrements or improves human performance, thus increasing or decreasing the likelihood of human error respectively, [42]. While there are many other PSFs that can affect human behaviour – for instance training, experience, competence, time available, workload, job design, manning, ergonomics of the equipment and procedures - these are excluded from the models as they are not affected by exposure to GDFs. All the excluded PSFs are implicitly assumed to remain constant within the model.

Other potentially relevant factors, which are not considered, are the long terms effects of GDFs on the crew performance. For example, we do not consider the hearing loss due to long-term noise exposure either individually or in combination with other GDFs effects. In practical terms only the effect of GDFs-affected human performance on the possible occurrence of collision and grounding in combination with the safety critical task (SCT) being performed are considered.

In the models the inputs and outputs are predetermined. The GDFs form the three inputs: ship motion, noise and whole body vibration (WBV). The unwanted outcomes form an output: the probability of an accident. The latter is chosen as a measure of human performance in SCT related to accident avoidance; since it is commonly accepted and widely used metric in the maritime domain thus it could be easily implemented in the RBSD process.

In reality, crew exposure to GDFs is likely to result in a plethora of effects on human performance and subsequent outcomes. However, to remain within the scope of our study the causal representation is limited to describing only those mechanisms that can describe the relationships between the predetermined inputs and outputs, as depicted in Fig. 2.

Two main paths linking GDFs exposure to human behaviour have been identified:

- Path 1: Stressor effects. Exposure to a GDF acts as a stressor and can affect the perceptual, cognitive and physical capabilities of an individual (e.g. attention management), which can subsequently impair the performance of the individual (i.e. the actual behaviour produced).
- Path 2: Physical effects. Exposure to a GDF can have specific and direct effects on the behaviour produced. For example, Ship motion can result in Motion-Induced Interruptions (MII). MII does not affect the underlying human capabilities of balance or fine motor control, but it exceeds the ability of the human to compensate and produce the intended behaviour. Similarly, WBV can directly impact the actual behaviour produced.

These two paths show how GDFs exposure affects human behaviour, which in turn influences the performance of SCT. It is the outcomes of an individual's actions and behaviour that determine the success or failure of a SCT. Insufficient performance of the SCT creates an antecedent for the unwanted outcome. The SCT are associated with:

- Maintaining safe vessel navigation thus avoiding collision or grounding.
- Proper maintenance of technical equipment of a ship, required for performing accident evasive action.

However, insufficient task performance alone does not determine

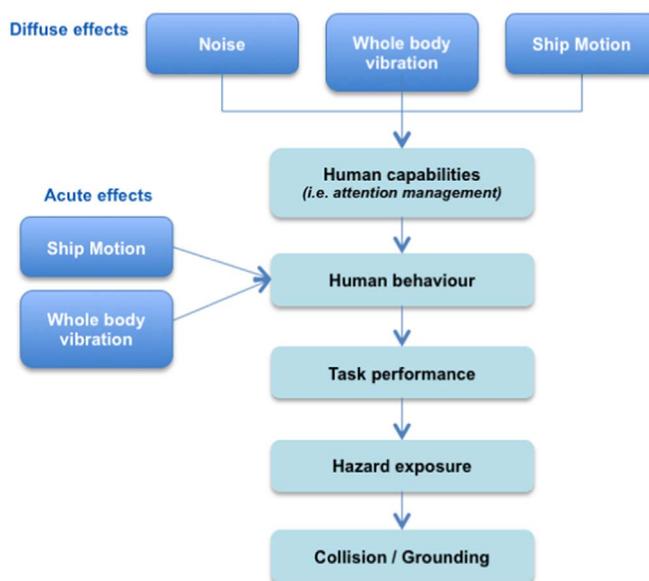


Fig. 2. A causal chain describing the relationship between crew, GDF exposure and unwanted outcomes, [43].

whether or not a collision or grounding occurs; the vessel must also be exposed to the collision or grounding hazard, as follows:

- For a collision to occur, another vessel must be on a collision course.
- For a grounding to occur, the ship must be in shallow water.

This causal mechanism makes the following assumptions:

- While we recognise that individuals have differing cognitive and physical abilities, it is assumed that all individuals have the same basic set of capabilities (i.e. all individuals can manage their attention, irrespective of the extent of this capability).
- Human behaviour is influenced by diffuse and acute effects of GDFs exposure as represented by the paths in Fig. 2.
- The crew perform SCT related to collision and grounding and tasks are appropriate, processes and procedures are optimised, and are undertaken by a competent operator.
- SCT must be performed correctly to maintain safe vessel operation.
- In case of bridge team the SCT manage the exposure of the vessel to the collision and grounding hazard.
- In case of engine room team the SCT manage the ability of a technical systems of a vessel to respond as requested by the bridge team.
- While it is recognized that interaction effects between GDFs within each pathway are likely to exist, these are excluded from the models, as the literature does not provide any information describing this interaction. For the review of the relevant literature the reader is referred to our earlier work [26,27].

3. Models development

3.1. Background knowledge related to human performance affected by the GDFs

It was found that the data on the specific GDF effects of ship motion (with the exception of motion induced interruption MII), noise, WBV on human performance are sparse and in many, but not all, cases generated under very specific, often non-marine, conditions. Data shows that there is certainly evidence for GDFs having some effect on human performance. However, the direct effects of GDF exposure on human performance tend to be weak, whereas secondary effects acting through another mechanism (e.g. fatigue, Motion Induced Sickness - MIS) tend to be stronger and more pervasive, [44,45]. Specifically, there are some data that describe the:

- Impact of GDFs on specific human capabilities, [46].
- Impact of GDFs on specific human behaviours, [47].
- Impact of errors on task performance, [48].

However, there is very little data about the link between the following components:

- Degraded human capabilities and collision or grounding related performance.
- Degraded task performance and exposure to the collision/grounding hazard

Fig. 3 demonstrates the links in the causal chain for which some quantitative data are available (in green) and the links for which there is no data (in red). For a summary of available literature the reader is referred to our earlier work, [27]. In addition to this gap, a given level of exposure to GDFs of certain intensity or duration may not affect all individuals equally. For example, while a given frequency and amplitude of ship motion may be generally MIS-inducing, individual experiences may range from significant nausea to no negative affects whatsoever, depending on their underlying susceptibility to MIS and

the degree to which they have acclimatized. Moreover, with the possible exception of secondary effects on human performance caused by fatigue, attributable to sleep disruption, a holistic view could not readily be derived directly from the individual findings. As such, the relevant theoretical models available in the scientific literature guided our approach.

The approach taken here to describe a mechanism that accounts for the impact of stressors on human performance has been based on the principles of attention management. It combines the principles from three theoretical models:

- Dynamic Adaptability Model (DAM), [29].
- Cognitive Control Model (CCM), [30].
- Malleable Attentional Resources Theory (MART), [31].

Under the DAM paradigm, GDFs are seen as types of physical stressor that affect human capabilities associated with maintaining a desired level of task performance either directly or indirectly (e.g. via fatigue). When exposed to GDFs, CCM describes humans compensating through the effortful direction of more cognitive resources at the task, typically at the cost of performance in other areas. Despite the sophisticated, and potentially subconscious, strategies humans have at their disposal, there is a limit to how much an individual can compensate without experiencing degradation in primary or secondary task performance. In addition, the extent to which human can compensate for task demands is not fixed. MART describes this compensatory capability changing as a function of task demands and associated arousal an individual experiences – attentional resources available vary as a function of load. When humans are in a state of under-load (i.e. bored) their pool of attentional resources is relatively small and will increase proportionately with the demands placed on them. However, there is a limit to how much the pool of attentional resources can grow. When task demands exceed the pool of attentional resources available (either transiently or when the upper attentional resource limit is exceeded), performance can breakdown and errors may be made.

Generally, task performance is only expected to degrade and become insufficient when compensatory mechanisms have failed. However, the literature does not allow prediction of how and when (chronologically) an operator would fail, under what conditions of GDF exposure, and what the specific effect on behaviour (i.e. type of error) would be.

In the models presented here, the main element is the performance of the SCT by engineers and a bridge team. The SCT of the engine team

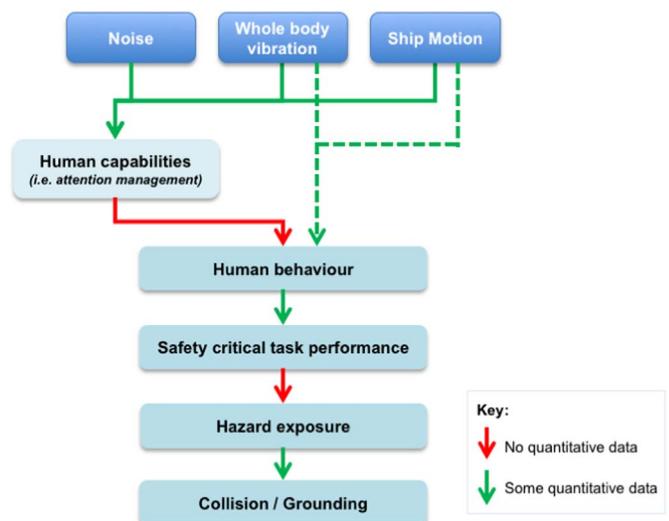


Fig. 3. Supporting data for links in the causal chain, [43].

is to maintain the ship systems, ensuring that they function properly when needed. This is a simplified description of the effect of GDFs on technical failure, through the mediating agent of crewmember. The SCT of the bridge team is to perform tasks associated with an accident evasive action. These tasks are complex and distributed in time, thus can be decomposed into three major phases, [5]: detection (D), assessment (A), action (Act). These three phases (DAAct) reflect the basic cognitive functions of observation, interpretation and planning, and execution, [49,50].

The performance of various tasks is governed by attention management, which is the supervisory human capability that directs, allocates and regulates the attentional resources required for the tasks. This high-level supervisory capability manages lower-level tasks such as perception, cognition, decision-making, memory, fine motor control and locomotion.

To evaluate the effect of GDF exposure on human performance the approach based on attention management theory is taken. Therein the effect of GDFs is represented as a stressor that sits either above or below the threshold of attentional capacity for any given SCT. If the stressor exceeds the attentional capacity then the attention management is degraded, whereas no negative effect is expected if the stressor can be managed within the available attentional capacity.

Representing ship motion, noise and WBV GDFs as stressors interacting with an individual's attention management capabilities provides an evidence-based mechanism for human performance that has been used to develop the models presented here.

3.2. Aggregation of the background knowledge into a model

The GDFs may have either an acute or diffuse effect. In case of an acute effect, the threshold refers to the GDFs level at which an individual may be unable to physically compensate for GDF exposure and perform actions as intended. In case of diffuse effect, the threshold refers to the amount of motion, vibration or noise an individual can endure before it acts as a stressor (with a corresponding stress response). The exact value of the threshold will vary between individuals and it is dependent upon previous experience, exposure duration and sensitivity. If the thresholds are not exceeded, then the attention of a crew-member is not affected, otherwise the attention management capability is degraded. It is not yet known what underlying psychological factors set an individual's baseline tolerance to GDF exposure. In reality, the overall impact of stressor exposure is likely to be determined by an individual's personal threshold for feeling the effects of stressor exposure, and the effectiveness of the strategies they adopt for managing mental resources to preserve task performance (e.g. task prioritisation and shedding).

Representing GDF exposure effects on safety behaviour via the attention management path provides a structure compatible with the introduction of an Error Producing Conditions (EPC) using the method called NARA,¹ which belongs to the latest (third) generation of the human reliability assessment (HRA) methods, and pertains to the nuclear field, [51].

The models begin with the discretization of the GDFs into the classes, which reflect the available background knowledge in the field. For the review of the effects of GDFs on humans, the reader is referred to earlier work, [27,28]. Based on that discretization, a type of an effect that the GDF has on a crewmember is determined, namely diffuse or acute along with the thresholds.

The diffuse effect of motion, vibration and noise, as well as the acute

¹ In recent years, the limitations and shortcomings of the second generation HRA methods have led to further developments related to the improvement of pre-existing methods. The only method now defined as third generation is nuclear action reliability assessment (NARA) and is, in fact, an advanced version of HEART for the nuclear field. The shortcomings in the second generation, have been the starting point of HRA experts for new research and improvement of existing methods, [80].

effect of the later is directly affecting the human capabilities of a bridge crew and the engine team, described by the concept of attention management. The latter if reduced, increases the probability of human error while performing a given type of SCT. If the attention management remains unaffected, the probability of error stays at its baseline level, as obtained from NARA.

Acute effects of all GDFs but noise don't affect attention management capability (AMC), but they directly degrade the physical ability of a crewmember. However, in case of the engine team, this effect is not anticipated, since we assume that the preventive maintenance on ships systems is not carried out if the levels of GDFs are above thresholds. This is in line with most of the operations guidelines of ships, where crews abstain from certain tasks if the weather conditions do not allow for safe performance of these tasks.

The probability of an error while performing SCT by the bridge team (detecting, assessing and acting along with communicating within the bridge team) together with the probability of technical failure, due to erroneous maintenance of critical ship systems by the engine team, yields the probability of not making proper and effective evasive action while on a course leading to an accident.

If the other ship on the collision course does not take appropriate action either, the collision is inevitable. In case of grounding accident, the lack of action from the ship on a grounding course results in the accident.

The models are presented in Figs. 4 and 5, and elaborated in details in the following sections.

3.3. Quantification of the GDFs and their effect on human capabilities

In the presented models the variables associated with the GDFs are as follows:

- noise (dB),
- vibrations (Hz),
- motion expressed by the accelerations in horizontal and vertical planes (in RMS g), as well the frequency of vertical motion (in Hz).

The variables are discretised according to the best available background knowledge, summarised in earlier work see [27,28], to reflect various effects that the GDFs alone, or with combination, may have on a person.

In the models presented here we assign the equal a priori probabilities for the states of GDFs. However, if the models are used to quantify the human performance associated with a specific ship design, these probabilities are updated accordingly to reflect the anticipated behaviour of the designs under evaluation. The discretization and a priori probabilities for the GDFs are gathered in tables presented in Appendix A (Tables A8–A12).

In the models the GDFs affect a person in a binary manner, either they act as a stressor, which degrades the human capability thus human performance, or they do not have any effect whatsoever. For the GDFs to become an active stressor their values need to be above threshold. The probabilities of GDFs sitting below or above a threshold, given set of GDFs, are evaluated by the conditional probability tables presented in the Appendix A (Tables A13–A20). For the summary of relevant background knowledge adopted to develop these tables see [27,28].

The quantification of the effect of GDFs stressors on the capabilities of human, through the concept of the AMC, is presented in Table 1. If the stressors are active (the GDF sits above the threshold), the AMC is degraded with the probability of 1. The analogical assumption is made for the situation where the stressors are inactive.

3.4. Quantification of the human behaviour in the presence of GDFs

3.4.1. Detection, assessment and action

Based on the earlier research, [5,49,50], and rational thought a

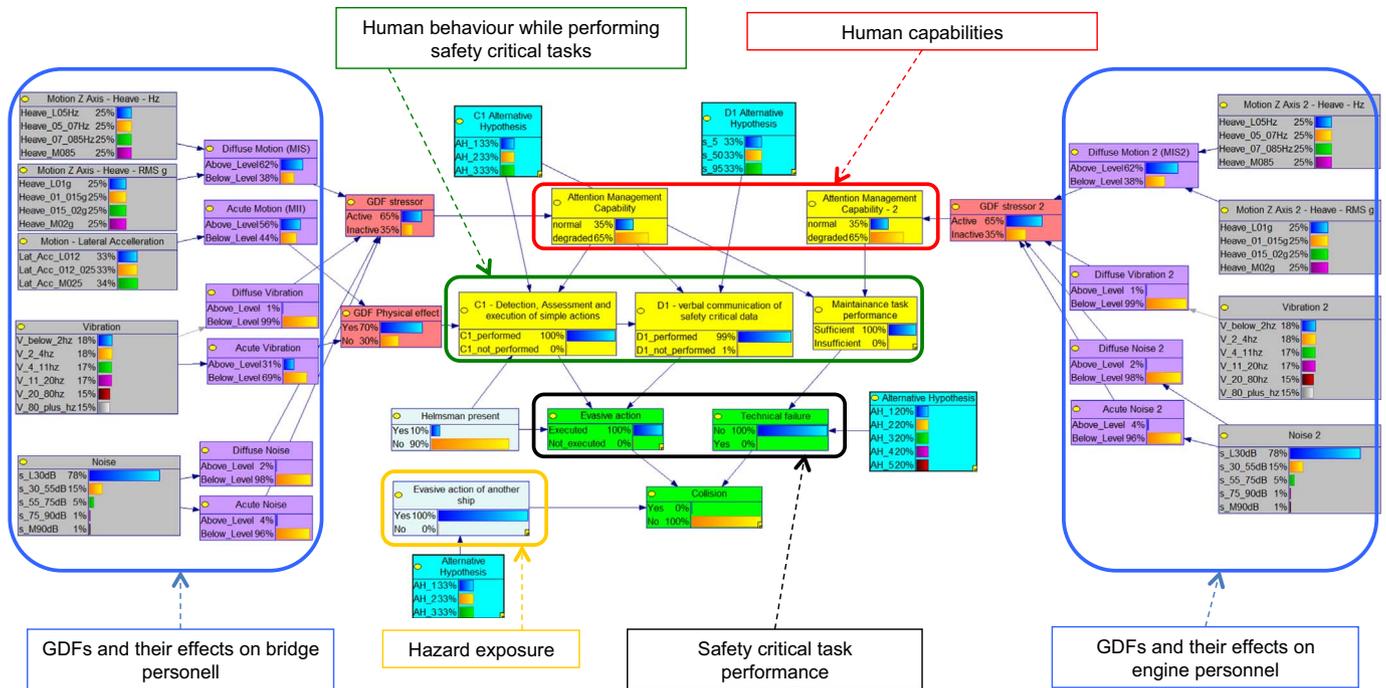


Fig. 4. A model for ship-ship collision accounting for the effect of GDFs.

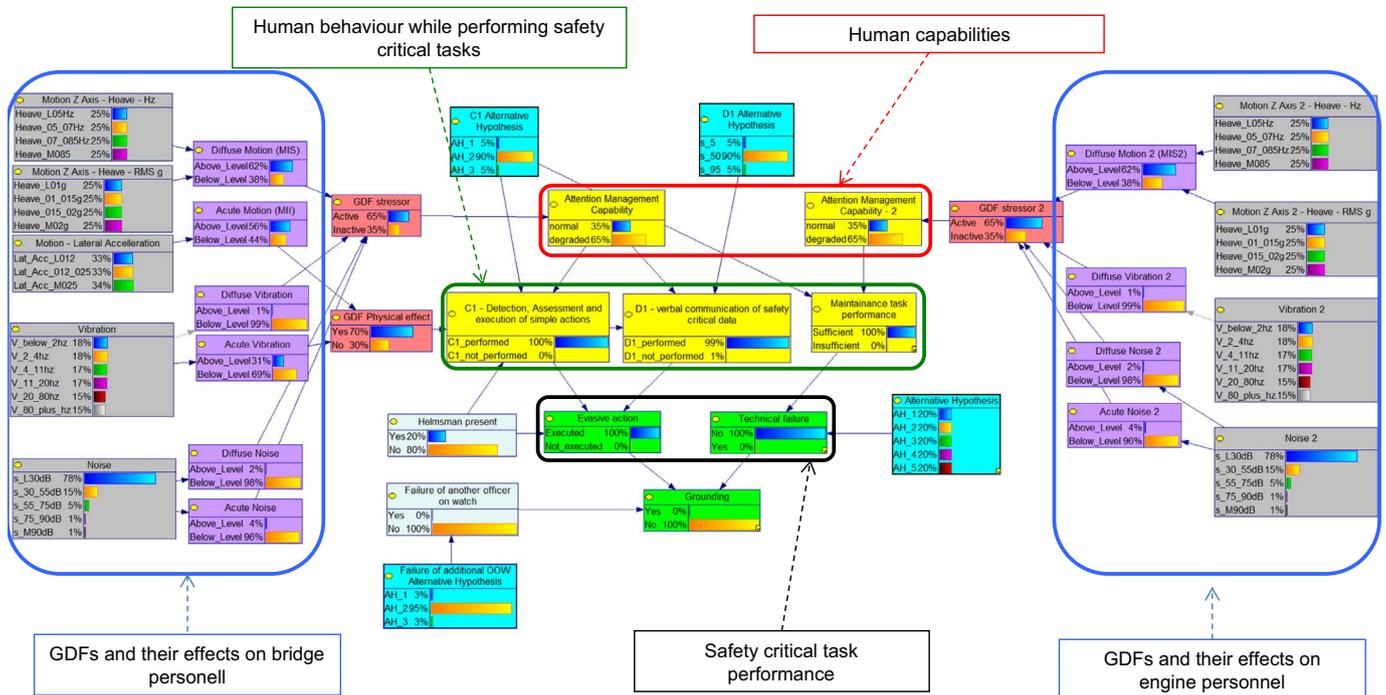


Fig. 5. A model for ship grounding accounting for the effect of GDFs.

Table 1

Probability table of the node Attention Management Capability and Attention Management Capability 2 (Parent node for the former *GDF stressor* and for the latter *GDF stressor technical*).

Attention Management Capability - AMC	Probability	
	Active	Inactive
<i>GDF stressor</i>		
Normal	0	1
Degraded	1	0

sequence of SCT related to accident avoidance can be broken down into three major phases: detection, assessment, action (DAAct). These are linked with the basic cognitive functions of human such as observation, interpretation and planning, and execution. In the models presented here the performance of a human with respect to these functions is modelled with the use of NARA guidelines and the generic task types (GTT) that it offers. To limit the complexity of the models, a single GTT was sought to represent all relevant navigational tasks performed by the OOW that are important in executing collision avoidance and

navigation in shallow waters. The GTT that is most analogous is, [51]:

Task C1 – Simple response to alarms/indications providing clear indication of situation (Simple diagnosis required) Response might be direct execution of simple actions or initiating other actions separately assessed. (Nominal HEP = 0.0005)

This task refers to the triplet DAAct, and it determines whether or not the SCT of detection, assessment and actions is performed – which implies that it is also sufficient. This is an all-encompassing definition including whatever tasks are required to maintain situational awareness and to respond appropriately to avoid collision or grounding. The probability of not performing DAAct is calculated based on the integration of three nodes: *Helmsman present*, *GDF Physical Effect* and *Attention Management Capability (AMC)*, as shown in Table B21 and depicted in Fig. 4.

In the case of not degraded AMC ($AMC=z_I$) and the absence of helmsman ($H=h_O$) we assume very weak physical effect of the GDFs ($PE=x_I$) on not performing the DAAct ($DAAct=\theta_O$), as follows:

$$p(DAAct = \theta_O | (H = h_O, AMC = z_I, PE = x_I)) = 0.001, \quad (1)$$

However, this effect is represented as being more significant in combination with a degraded AMC ($AMC=z_O$), as follows:

$$p(DAAct = \theta_O | (H = h_O, AMC = z_O, PE = x_I)) = 0.0011, \quad (2)$$

Thus representing an additional drain on cognitive resources compensating for physical task disruption. These values were estimated using judgement. To set the probability of not performing DAAct when a person is not affected by GDF exposure ($PE=x_O$), to reflect a baseline error rate, the NARA GTT value for Task C1 of 0.0005 is used, as follows:

$$p(DAAct = \theta_O | (H = h_O, AMC = z_I, PE = x_O)) = 0.0005, \quad (3)$$

NARA categorises the factors that negatively influence human performance as one of eighteen Error Producing Conditions (EPCs), see Table B28. The EPC that best represented the causal mechanism from GDFs exposure to human performance was EPC No. 15 ‘Poor Environment’. This EPC represents the stressor effect of GDF exposure on attention management capability. The potential strength of effect of this EPC was set using the Assessed Proportion of Affect (APOA) variable. The APOA level was set based on the application of the NARA methodology to subjectively determine an appropriate value, nominally between 0 (no effect) and 1 (maximum effect). However, based on the guidance available for NARA, it was decided to cap the maximum APOA associated with the EPC to 0.1.

Then using NARA guidelines, given exposure to error producing conditions (EPC) No. 15, the probability of human error (HEP) in the above mentioned situation of 0.0006 is calculated, see Table B21. This HEP corresponds to the probability of not performing DAAct, as follows:

$$p(DAAct = \theta_O | (H = h_O, AMC = z_O, PE = x_O)) = 0.0006, \quad (4)$$

The NARA calculation allows inclusion of multiple EPCs and an Extended Time Factor (ETF). In the models presented here GDFs are represented using only one EPC and there is little justification to include the ETF. Thus, the HEP is calculated based on the following formula:

$$HEP = GTT \cdot [(EPC - 1) \cdot (APOA + 1)] \quad (5)$$

3.4.2. Verbal communication on safety critical data

If evasive manoeuvres are performed with the presence of helmsman, which requires appropriate communication of information between the OOW and helmsman, the node *D1 - verbal communication of safety critical data* is evaluated, as presented in Table B22. This node determines whether or not the verbal communication of safety

critical data to the helmsman required to avoid accident is sufficient. It relates to the NARA GTT No. D1 of the same name. This node is passed to its descendant called *Evasive Action*, where it is used to quantify the probability of this action differentiating between the situation where the helmsman is present or not. The NARA GTT value for Task D1 of 0.006 is used to set the probability of *insufficient performance* ($D_I=d_O$) unaffected by GDF exposure ($AMC=z_I$) to reflect a baseline error rate.

$$p(D1 = d_O | (AMC = z_I, DAAct = \theta_I)) = 0.006, \quad (6)$$

The HEP of 0.0072 was calculated using NARA given exposure to EPC No. 15 representing the effect of GDF exposure via AMC ($AMC=z_O$), reflecting the following situation:

$$p(D1 = d_O | (AMC = z_O, DAAct = \theta_I)) = 0.0072, \quad (7)$$

It is assumed that *Verbal communication of safety critical data* is unaffected by the *GDF Physical Effect*.

3.4.3. Evasive action

Evasive action is assessed in the node *Evasive Action*, as shown in Table B23. It is assumed that if *Detection, assessment and execution of simple actions* and *Verbal communication of safety critical data*, where applicable, are performed then *Evasive Action* will be executed. If the helmsman is present then the probability of not executing the evasive action is capped at 0.0001, in line with the NARA Human Performance Limiting Value for ‘Actions taken by a team of operators’.

3.4.4. Technical failure

Technical failure node quantifies the probability of the relevant systems not functioning as a result of lack of maintenance or poor maintenance caused by the GDFs affecting the AMC of a crewmember responsible for the maintenance. This node was included in recognition of the importance of maintenance in sustaining the functionality of vessel equipment such that it performs as it is designed to. Errors during maintenance on systems that provide the manoeuvring capability of the vessel can limit the vessel’s response to control inputs associated with evasive action, hence affecting the probability of an unwanted outcome. *Technical failure* node determines whether or not maintenance actions performed on equipment that provides the vessel’s manoeuvring capability has been completed successfully, see Table B24.

The probability of insufficient *Maintenance Task Performance* ($MTP=m_O$) is calculated based on the *Attention Management Capability-2* node, which follows the principle presented in Table A19. To set the probability of insufficient *Maintenance Task Performance*, unaffected by GDF exposure, to reflect a baseline error rate, the NARA GTT value for Task C1 of 0.0005 is used, as follows:

$$p(MTP = m_O | AMC2 = z_I) = 0.0005, \quad (8)$$

The HEP of 0.0006 was calculated using NARA given exposure to EPC No. 15, representing the effect of GDF exposure via the *Attention Management Capability-2* node, as presented in Table B25.

$$p(MTP = m_O | AMC2 = z_O) = 0.0006, \quad (9)$$

The GDF Physical Effect on the probability of insufficient safety behaviour is not anticipated. This comes from an assumption about lack of preventive maintenance carried out if the levels of GDFs are above thresholds. This is in line with most of the operations guidelines of ships, where crews abstain from certain tasks if the weather conditions do not allow for safe performance of these tasks.

3.4.5. Evasive action of another ship

Evasive action of another ship is a node, which accounts for the behaviour of OOW on a ship that is encountered by the own vessel. The probability for this node is assigned based on NARA calculations, and subject to alternative hypothesis testing, as elaborated in Section 3.5.

Table 2
The qualitative assessment of evidential uncertainty for the models presented here, [58].

Model parameter	Justification for the evidential uncertainty score	Evidential uncertainty score
Maintenance Task Performance	This node represents the performance of navigation tasks critical in collision or grounding avoidance and provides a structure compatible with the introduction of a NARA GTT, potentially affected by EPC No. 15 via ‘Attention Management Capability’ and GDF Physical effects	Moderate
C1 - Detection, Assessment and execution of simple actions		
D1 - verbal communication of safety critical data	This node represents the communication of vessel manoeuvring instructions critical in collision or grounding avoidance with a helmsman present and introduction of a NARA GTT, potentially affected by EPC No. 15 via ‘Attention Management Capability’.	Moderate
Evasive action of another ship	The node represents the performance of navigation tasks critical in accident avoidance on board other ship. It is quantified based on NARA.	Moderate
Helmsman present	At the moment this node is quantified fully based on judgement. However, more detailed assessment is possible if needed, for instance by performing survey among shipping companies.	Moderate-Minor

3.4.6. Failure of another officer of the watch

This node exists in the model quantifying the performance of a navigator conducting SCT related to avoiding grounding accident. The node reflects the team-work nature of bridge navigation, which is a process, where dangerous situations are anticipated well before they arise. In case of collisions the time horizon for such anticipation is usually expressed in minutes,² however in case of groundings it can be in hours or even days. A navigator knowing the passage plan is aware of any areas that are potentially dangerous, especially if the ship course is not adjusted and a ship leaves the pre-planned and safe route, [52]. Moreover, before a new bridge watch commences the relieved officer reports to the one who takes over all the anticipated course alterations and dangers to navigation to be expected during his watch, thus increasing his situational awareness, [53].

Such practice is recognized in the grounding model by the node *Failure of another officer on the watch*. This node quantifies the probability that an officer leaving the bridge will not pass the relevant information on anticipated hazardous waters to the officer taking over.

The probability for this node is assigned based on NARA calculations, and subject to alternative hypothesis testing, as elaborated in Section 3.5.

3.4.7. An accident

The last nodes of the models quantify the probability of an accident, namely collision and grounding, for a given set of GDFs. However, the models do not look to estimate probability of accident per se. Instead they take it as a well-understood measure of human performance, to propose a way to reflect human performance positively, which can be maximized in ship design phase. The logic behind the nodes is presented in Tables B26 and B27.

3.5. Uncertainty assessment and treatment

The parameters of the models presented here are developed based on certain amount of background knowledge, which by no means is complete. This incompleteness is the primary source of uncertainty associated with the models and their parameters, [54,55]. Some elements of the model and their relations are characterised by larger amount of information, knowledge and understanding, thus have lower uncertainty, the others are known and/or understood less resulting in larger uncertainty.

The uncertainty assessment together with the sensitivity analysis of the models provides a valuable tool for screening the models for important variables, which are both uncertain and the models are

² In situations where two ships are approaching each other at parallel courses and very low relative speed, the time to the closest point of approach (TCPA), can be expressed in hours. This measure along with the closest distance between two ships encountering (CPA) is an example of the operational measures of safety during sea passages. However in most cases the time needed to solve potential collision encounters is not more than tens of minutes.

sensitive to. By ranking the variables in the models by their importance, we define the set of variables affecting the most the credibility of the models. Finally, with respect to the variables, which are important, the uncertainty treatment is applied, as described in the following sections.

3.5.1. Sensitivity analysis

A sensitivity-value approach presented by Coupé and van der Gaag [56] is applied here. 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. Based on the findings from the sensitivity analysis of the models, the following can be concluded:

- The models are highly sensitive to the following parameters: *Maintenance Task Performance, C1 – Detection, Assessment and execution of simple actions and D1 – verbal communication of safety critical data.*
- The models are moderately sensitive to *Evasive action of another ship and Helmsman present*. However, the effect that these parameters have on the output is significantly lower than the effects of *C1* and *D1*, as specified above.
- The remaining nodes have very low sensitivity values, meaning that their effects on the models outputs are rather minor.

3.5.2. Evidential uncertainty assessment

For the most sensitive model parameters the evidential uncertainty assessment is carried out, and the results are presented in Table 2. To rank the uncertainty we apply the following qualitative scoring system as introduced in [57]:

3.5.2.1. Significant uncertainty. All of the following conditions are met:

- The phenomena involved are not well understood; models are non-existent or known/believed to give poor predictions.
- The assumptions made represent strong simplifications.
- Data are not available, or are unreliable.
- There is lack of agreement/consensus among experts.

3.5.2.2. Minor uncertainty. All of the following conditions are met:

- The phenomena involved are well understood; the models used are known to give predictions with the required accuracy.
- The assumptions made are seen as very reasonable.
- Much reliable data are available.
- There is broad agreement among experts.

Table 3
The qualitative assessment of model parameters importance for the models presented here [58].

Model parameter	Evidential uncertainty score		Sensitivity score	Importance score
Maintenance Task Performance	Moderate	High		High
C1 - Detection, Assessment and execution of simple actions	Moderate	High		High
D1 - verbal communication of safety critical data	Moderate	High		High
Evasive action of another ship	Moderate	Moderate		Moderate
Helmsman present	Moderate-Minor	Moderate		Moderate

3.5.2.3. *Moderate uncertainty.* Conditions between those characterising significant and minor uncertainty, e.g.:

- The phenomena involved are well understood, but the models used are considered simple/crude.
- Some reliable data are available.

Subsequently, by combining the results of sensitive and uncertainty assessment, the parameter importance ranking is carried out, see Table 3. Three parameters have high importance score, namely *Maintenance Task Performance*, *C1-Detection, Assessment and Execution of Simple Action*, *D1 – verbal communication on safety critical data*. For those three elements the uncertainty treatment is carried out as presented in the following section.

3.5.3. *Uncertainty treatment*

The uncertainty associated with the model parameters, which are believed to be to some extent known, is expressed through the probabilities. For example, the state of a specific GDF associated with a given ship design, say noise, can be described by a probability density function that reflects the anticipated variability of this parameter for the given design across a set of operational conditions.

In a similar way we can judge the probability of a helmsman being present during a critical situation. This probability can be based on the analysis of shipborne operations and the way how the bridge watch is organized and conducted, knowing that most of the day-time the bridge is manned by one person only, and a watchman is present during night shifts and when navigating in coastal waters.

However, the majority of the uncertainty in the models comes from the limited knowledge related to human performance given the presence of stressors (GDFs). Despite the extensive amount of data that NARA is built upon, it is very difficult to point to a given probability as the probability of erroneous behaviour of a human in certain situation.

The probability of human error (HEP) is estimated based on NARA methodology, for a set of factors and conditions, as presented in

Table 4
The lower and upper probabilities for the parameters obtained from NARA, [51].

GTT and nominal probability	P(GTT)	P̄(GTT)	EPC1	APOA1			HEP	HEP	HEP	
C1.	5E-4	3E-4	1E-2	15-Poor Environment	3	Low	0,1	6E-4	3.6E-4	1.2E-2
	5E-4	3E-4	1E-2	15-Poor Environment	3	High (N/A)	0,8	1.3E-3		
	5E-4	3E-4	1E-2	N/A	0	NIL	0	5E-4		
D1.	6E-3	3E-3	1E-2	15 - Poor Environment	3	Low	0,1	7.2E-3	3.6E-3	1.2E-2
	6E-3	3E-3	1E-2	15 - Poor Environment	3	High (N/A)	1	1.8E-2		
	6E-3	3E-3	1E-2	N/A	0	NIL	0	6E-3	3E-3	1E-2

C1: Simple response to alarms/indications providing clear indication of situation (Simple diagnosis required) Response might be direct execution of simple actions or initiating other actions separately assessed.

D1: Verbal communication of safety critical data.

Table 4. However, our background knowledge on the link between human capability and human behaviour is heavily limited, as depicted in Fig. 3, and we are unable to gain new information at this point. Acknowledging that fact, we have to select an appropriate way of dealing with this type of uncertainty, trying to improve the model performance and results communication, [59]. We do not expect uncertainty reduction, since our limited knowledge will not improve, but we would expect some bounds for the results to fall within. Since, the NARA itself offers the intervals for the probability to account for the anticipated uncertainty; this seems to be a rational way out in our case.

There exist several approaches to the uncertainty treatment, see for example [57,60–67]. One of those, suitable for the treatment of epistemic uncertainty is the use of imprecise probabilities, in this case the intervals, along with the alternative hypotheses testing (AHT), [68,69].

The AHT, in principle does not reduce the uncertainty but helps to determine its bounds, which is relevant for decision making process. Even though the results obtained are not more accurate than the results without using AHT, the calculated bounds provide a warning for a decision-maker about the quality of a given parameter, which is missing in case of point estimates. In our case, the AHT is performed for the variables that score high in importance ranking. These are associated with the performance of a human affected by external stressors. The human error probability (HEP) is calculated for a given Generic Task Type (GTT) given set of stressors described by, so called, Error Producing Conditions (EPC).

The probabilities of nominal HEP for a given GTT are defined as intervals, whereas the EPCs are chosen by the analysts. This provides several combinations leading to several coexisting alternatives ($AH_1, AH_2...AH_n$). Due to lack of knowledge we cannot point to any specify hypothesis to be true for the given case, however we believe that we can judge their probabilities of occurrence, using subjective probabilities ($p(AH_1)+p(AH_2)+...+p(AH_n)=1$), [70,71]. By doing that we obtain a set of HEPs rather than a single value HEP.

The background knowledge and the associated uncertainty are propagated through the model, and the output is produced. The latter is a two-state variable $Z=(z1, z2)$, where $z1=accident$; $z2=no_accident$, and the quantity of interest is the probability of the accident given the set on input variables (X), as follows $p(Z=z1|X)$.

Table 5
The probabilities of collision and grounding obtained in the course of extreme conditions tests.

	The probability of collision		The probability of grounding	
	An absolute value	Change compared to the baseline	An absolute value	Change compared to the baseline
All active	1,38E-06	+10%	3,58E-04	+16%
Base line	1,26E-06	–	3,06E-04	–
All inactive	7,01E-07	–45%	2,11E-04	–31%

The following parameters of models need to be described with the use of imprecise probabilities, characterised by the lower and upper bounds:

1. C1 – Detection, assessment and execution of simple actions.
2. D1 – Verbal communication of safety critical data.
3. Maintenance task performance.
4. Technical failure.
5. Evasive action of another ship.
6. Failure of another officer of the watch.

All the items but 4 are based on the same paradigm and utilize the same upper and lower bounds for factors specified by NARA. The item 4 provides the spectrum of failure probabilities for a technical system, for normal and degraded maintenance regime, based on the historical data and experts’ judgement. The values of the imprecise probabilities adopted for the items 1–3 and 5–6 are gathered in Table 4. The probability of evasive action of another ship and the failure of another officer of the watch correspond to the probabilities of C1.

The probability of technical failure – item 4 – is estimated as an interval - $\Delta p(\text{techn_failure})$, where the upper - $\bar{p}(\text{techfailure})$ - and lower - $\underline{p}(\text{technfailure})$ – probabilities are obtained from the literature,³ [72,73]. This interval describes the probability of technical failure in normal operation, where the maintenance tasks are taken properly ($MTP=m1$), as follows $\Delta p(\text{techn_failure}|MTP=m1)$.

However, there is no evidence for the probability of technical failure during erroneous maintenance, which results from the reduced human performance. In such situation the following three interpretations, referred to as alternative hypothesis (AH), may hold.

First it is intuitive to expect the probability of technical failure to increase, compared to normal maintenance. The most straightforward way to model that is to assume, that in case of improper maintenance ($MTP=m0$) the probability of technical failure equals 1.

$$AH1 : p(\text{techn_failure}|MTP = m_0) = 1, \tag{10}$$

Second, in case of erroneous action, the redundancy of the system may prevent technical failure to affect the operations of the safety critical systems, thus the probability of technical failure will remain the same as in normal operations:

$$AH2 : p(\text{techn_failure}|MTP = m_0) = \Delta p(\text{techn_failure}), \tag{11}$$

Third, in light of our limited knowledge on what will happen in case of erroneous maintenance in terms of the availability of safety critical technical systems of a ship, we can assume that both failure and lack of it are equally likely:

$$AH3 : p(\text{techn_failure}|MTP = m_0) = 0.5, \tag{12}$$

To account for three interpretations, we consider each of these as an

³ In [72] the probability of a technical failure in a critical situation, yields 4.5E-6, based on accident statistics. In [73] the probability of technical failure per hour of ship operation, elicited from the experts yields 3E-4.

Table 6
The results of comparison of two nomologically proximate models.

States of selected nodes in the risk models	DNV model		The models presented here	
	Prob. of collision	Relative change	Prob. of collision	Relative change
Increase performance performance=excellent/ AMC= normal	8,1E-06	–5%	1,0E-06	–20%
Base line	8,60E-06	0	1,26E-06	0
Deteriorate performance performance=poor/ AMC=degraded	9,5E-06	+10%	1,4E-06	+11%

Table 7
The summary of validity tests.

Validity criteria	Score
Face	High
Content	Moderate
Predictive	Moderate-High
Concurrent	Moderate

alternative hypothesis and we assign a weight to each, corresponding to our belief on a given variable becoming true. At the moment there is no support for giving preference to any of these, we assign equal weights for the hypothesis, on the probability of technical failure given insufficient maintenance as follows:

$$p(AH1)=p(AH2)=p(AH3) = 1/3, \tag{13}$$

However, in light of new evidences these weights and probabilities may be easily updated in the models.

4. Validation of the models

This section shows the results of validation framework applied to the models described in the earlier sections. The validation framework assesses the plausibility of the model as a tool for serving its envisaged functions: (i) to convey an argumentation based on available evidence, (ii) to discriminate different ship designs.

4.1. Translation validity

4.1.1. Face validity

The face validity of the models is drawn from the evidence base that drove the development of the model. The major components that make up the model of the attention management mechanism and were derived from the literature, as described earlier in this paper.

The input node thresholds for effect on human performance were derived from the literature, however the probabilities were set by a combination of expert judgement and human reliability assessment (HRA) methods. The NARA guidelines selected is evidence-based to the extent it utilises the CORE-DATA set to set basic human error probabilities that are derived from real-world human reliability data. However, significant expert judgement is required to select which factors are appropriate for the context and to determine the magnitude of their effect on human performance in the context. Despite the adoption of HRA methods in a number of domains (e.g. nuclear, oil & gas), these subjective judgements required by the method remain subject to potential inter-rater variability.

There are very few models inhabiting the modelling space presented here, hence it is difficult to elaborate on their consistency with the parameters of these models. Some elements existing in the models

correspond to the expert driven risk models for collision and grounding presented in the literature. Structurally, the concept of degraded and normal attention and resulted human behaviour are the aspects the models have in common with a model proposed by DNV, [13,74]. However, the relationship of the attention management construct to other elements in the model is unique to the models presented here.

Overall, the face validity of the models presented here is high.

4.1.2. Content validity

The development of the models was based on scientific literature as far as possible. The pre-selected GDFs of ship motion, noise, WBV are a subset of the potential performing-shaping factors that could influence the outcomes of collision or grounding. Other factors can be imagined that are influenced by vessel design, which have a greater impact on human performance, however the three types of GDFs were chosen arbitrarily.

The states for the GDF input variables of ship motion, noise and vibration were derived from the literature around thresholds of effect that, despite incomplete data, can be justified with some caveats.

Other states for nodes within the model represent a simplification due to incomplete knowledge (attention management being represented by only two states: normal or degraded), or to perform a function within the model (switching GDF stressor effects and GDF physical effects on or off). These binary nodes are understood to have low content validity.

Nodes outside the GDF input variables fail the test of dimensional consistency as they do reflect states either above or below a level, but cannot be set at the level itself. Hence these reduce the content validity of the model.

The presented models are developed to the limits of the relevant literature and expert knowledge. The integration of HRA is relevant here. Despite HRAs general limitations as a method in terms of reliability and its application here being outside its normal use, it was necessary as a means to generate probabilistic bounds for the nodes within the risk model.

Overall, the content validity of the models is moderate, being compromised by the limitations imposed by the limited scientific knowledge in the domain.

4.2. Criterion related validity

4.2.1. Concurrent validity

Concurrent validity (CV) is tested in two-fold. First, by comparing the elements of the models with another model developed for a similar purpose. Second by comparing the models with the trends existing in the results from the experiments, which were designed for this specific purpose, conducted with the use of full mission bridge simulator as reported in [75].

The former is complicated, since there is no other model that would link the GDFs with the human task performance. The external models that were used for this validation exercises, were developed for specific purposes, which substantially differ from the purpose of the models presented here. Therefore, the inter-models comparison is feasible only with respect to selected elements of the models.

For instance, the elements describing human behaviour in a given situation can be compared with respect to the paradigms governing these, discretization level adopted or input-output relation. In some case qualitative comparison is feasible; in other quantitative evaluation is the only option.

There are several models that can be deemed relevant for this step of validation, for example [5,6,76–78]. However only one by DNV, is found suitable for the concurrent validity, [13,74].

4.2.1.1. Comparison with other relevant models. A model by DNV estimates the probability of collision, accounting for the navigational parameters, safety culture, personnel factors, management factors, technical reliability and other vigilance. The model assumes the failure of a navigator along the path of detection-assessment-action for an accident to happen. As a modelling technique BBN is adopted and experts' judgment is used as a primary source of information to parameterize the model. The model was developed to measure the effects of bridge layout on the collision and grounding probability through the estimation of performance of a navigator. The performance affects the causal path of: detection, assessment and action, moreover, the probability of technical failure of ship machinery is accounted for, likewise in the models described here.

Looking into the core of DNV model, where the effect of stressors on the performance of a navigator is estimated, we can say that our models and the DNV model can be compared at a generic level. The basic causal relations: stressor – attention of a person – detection – assessment – action is retained in both models, the logic behind their structure can be seen as similar.

The structure of our models is based on the concept of Attention Management and the parameters are found through the HRA methods and experts' judgments. The DNV model to a large extent is based on experts' judgment and historical data, therefore the structure and parameters of these two models can hardly be comparable.

Despite that difference, the validity of our models shall not be seen compromised. Since the purposes of the two compared models are substantially different, their input nodes and CPTs are not expected to match fully.

4.2.1.2. Comparison with simulator studies. By comparing the effect of GDFs on the performance of a navigator with the results obtained from experiments with the use of full mission ship bridge simulator, one can evaluate the level of concurrent validity of the relevant parts of the models. The results of simulator studies conducted for the purpose of validating the models shown here are reported in [75]. The performance of a navigator is measured in two-fold for each simulation run. First, two navigation measures are evaluated, i.e. normalised distance to the relevant target ship, and normalised track deviation. Second the instructor ratings on the rapidity/quality of mariners' responses are recorded. The studies account for the effect of noise on human performance. Due to technical limitations of the simulator the effect of motion and vibration could not be realistically modelled.

In the report authors claim that there are no significant effects of noise or task difficulty on either of the navigation measures. However, when the instructors' ratings of the rapidity/quality of mariners' responses are considered, a number of significant or marginally significant effects of noise and task difficulty emerge. The overall, global ratings of task performance by instructors showed both a significant effect of task difficulty (which validated the task difficulty manipulation within the experiment) and a significant effect of noise such that higher levels of noise were associated with poorer performance. The effects of noise were also either significant, or close to statistical significance, in the case of several specific actions within high difficulty scenarios.

This finding suggests that there might be an interaction between noise and task difficulty such that noise has little or no effect in situations where the task is easier, but might lead to impaired performance of a high difficulty task. These findings support, to some

extent, the modelling choices adopted here.

The overall concurrent validity of the risk model can be judged as moderate.

4.2.2. Predictive validity

Extreme conditions test is a case of predictive validity testing where the hypothetical model is set to extreme conditions where the behaviour of the model is more predictable, [79]. The results of these tests are presented in Table 5, where the measure of human performance is tabulated, expressed as the probabilities of collision and grounding. The results are shown for three different states of input variables. The base line refers to the input variables being in their states reflecting the average conditions on board ships. The all-active state refers to a situation where all the stressors are active and are affecting negatively the performance of a navigator. The all-inactive state refers to the situation where there is no stressor acting on a navigator.

4.2.2.1. Comparison of the results obtained from the models being in nomological proximity. There is only one model that is nomologically close to the model presented here, namely DNV risk model, **[13,74]. The proximity of these two models comes from the fact, that both measure the effect of performance shaping factors (PSFs) on the ability of a navigator to perform his main tasks in relation to safe navigation and collision avoidance action. However, the PSFs that each model takes into account are different, likewise the mechanisms governing that ability. In the model presented here navigator's ability is governed by the attention management capability, which is modelled by a variable of the same name. In DNV model this ability is modelled through a set of variables, where the one called *performance* seems to be the most comparable with the attention management in our model. Therefore, the effect of manipulation of these two variables on the explanatory variable is measured. The results of this test are presented in Table 6, where the probabilities of collision associated with the various states of the analysed variables are shown.

It is evident, that the behaviour of both models is comparable, when changing the states of the variables of interests between their states. In our model the variable *Attention Management Capabilities* has two states (*Normal*, *Degraded*), however in DNV model the variable *Performance* has four states (*Excellent*, *Standard*, *Poor*, *Not able to perform*). For comparison we took *Performance=Excellent* and *Performance=Poor*, these refer to *Normal* and *Degraded* attention management capability (AMC).

On the basis of the obtained scores of the validity criteria presented in Table 7, the overall predictive validity of the risk model can be judged as between moderate and high with the following observations:

- Moderate content validity informs that there is room for improvement when it comes to the parametrization of the models. This stems mainly from the lack of data about the analysed phenomena. In order to increase the score in this validation test, an extensive research is needed dedicated solely to the assessment of the effect of GDFs on performance of a crew member.
- Moderate concurrent validity reflects the actual state of the art in the analysed field. Due to lack of compatibility between the existing models and these presented here the direct comparison of these is not feasible. In some cases only elements of the models can be compared, but even then due to substantially different modelling paradigm behind the models the results of comparison are not satisfactory.

- Moderate – High predictive validity shows that the models developed behave as expected when tested. This means that the model ranks the designs appropriately for different GDF levels. In practice it means, that the exact numbers that the models provide may not be “correct” but the proper behaviour of the models means they can be used to distinguish different designs in a multi-objective optimization, which is the primary aim of the models.
- Also the sensitivity-uncertainty assessment allowed defining the most critical elements of the models and remedial actions were taken to improve the overall performance of the model. These comprise the alternative hypothesis with respect to the most critical element. This high score supports the adopted modelling techniques, which make it possible to carry out an extensive tests pertaining to the predictive validity.

5. Conclusion

This paper offers two novel models linking the effect of GDFs with human performance, which can be incorporated into a process of multi-objective ship design. Within the models the GDFs can be appropriately modified at the early stage of ship design process resulting in a design characterised by the highest performance of a human. All other factors affecting the human performance, but not belonging to this particular aspect of ship design, are omitted in the model (they are considered constant in all analysed design alternatives, thus having the same effect across designs).

The primary aim of the framework is to allow differentiation among various designs based on the criteria selected, which is human performance in safety critical tasks pertaining to accident avoidance. The results of validation process adopted here show that the presented models are valid for the given purpose, despite certain limitations and the paucity in data. These gaps in background knowledge lower the level of accuracy of the models, thus the models shall not be used to seek an accurate estimates of the measure of human performance (the probability of an accident). Instead the models can be used for the relative comparison of designs, which is affected to much lesser extent by the gaps in background knowledge.

The models may be used by naval architects, vessel designers, and vessel system designers as intended, provided access to Human Factor (HF) expertise is available to assist with application and interpretation. It is important to recognise the relevance of human factors input during its eventual application. HF provides the understanding of the complexities of human behaviour in operational settings, its interdependencies and interactions.

Acknowledgements

The authors appreciate the financial contributions of the EU, as this research was co-funded by the FAROS project (Project no. 314817, 01/10/12 – 01/10/15). The views expressed remain solely those of the authors.

Markus Porthin from VTT Finland is appreciated for his assistance in the early stage of the models development. The authors are also indebted to the anonymous reviewers and the editor whose comments improved the presentation and the content of an earlier version of the manuscript.

The probabilistic models introduced in this paper were created using the GeNie modeling environment developed at the Decision Systems Laboratory, University of Pittsburgh, available from <http://genie.sis.pitt.edu/>.

APPENDIX A. Quantification of the GDFs and their effect on human capabilities

(See Tables A8–A20).

Table A8

Probability table of the node Motion_Z_Axis_Heave – Hz.

Motion Z Axis – frequency (Hz)	Probability
Below 0.5	0.25
Between 0.5–0.7	0.25
Between 0.7–0.85	0.25
Above 0.85	0.25

Table A9

Probability table of the node Motion_Z_Axis_Heave – RMSg.

Motion Z Axis - acceleration (RMSg)	Probability
Below 0.1	0.25
Between 0.1–0.15	0.25
Between 0.15–0.2	0.25
Above 0.2	0.25

Table A10

Probability table of the node Motion - Lateral Acceleration.

Motion - lateral acceleration (RMS g)	Probability
Below 0.12	0.34
Between 0.12–0.25	0.33
Above 0.25	0.33

Table A11

Probability table of the node Vibration and Vibration 2.

Vibration – WBV (Hz)	Probability
Below 2	0.167
Between 2–4	0.167
Between 4–11	0.167
Between 11–20	0.167
Between 20–80	0.167
Above 80	0.165

Table A12

Probability table of the node Noise and Noise 2.

Noise (dB)	Probability
Below 30	0.2
Between 30–55	0.2
Between 55–75	0.2
Between 75–90	0.2
Above 90	0.2

Table A13

Probability table of the node Diffuse Motion (MIS) and Diffuse Motion 2 (MIS_2) (Parent nodes for the former: *Motion_Z_Axis_RMSg* and *Motion_Z_Axis_Hz*; Parent nodes for the latter: *Motion_Z_Axis_RMSg* 2 and *Motion_Z_Axis_Hz* 2).

Diffuse Motion (MIS)					Probability			
<i>Motion_Z_Axis_RMSg</i>	<i>Heave_L01g</i>				<i>Heave_01_015 g</i>			
<i>Motion_Z_Axis_Hz</i>	<i>Heave_L05hz</i>	<i>Heave_05_07hz</i>	<i>Heave_07_085hz</i>	<i>Heave_M085</i>	<i>Heave_L05hz</i>	<i>Heave_05_07hz</i>	<i>Heave_07_085hz</i>	<i>Heave_M085</i>
Above_Level	1	0	0	0	1	1	0	0
Below_Level	0	1	1	1	0	0	1	1

Diffuse Motion (MIS)					Probability			
<i>Motion_Z_Axis_RMSg</i>	<i>Heave_015_02g</i>				<i>Heave_M02g</i>			
<i>Motion_Z_Axis_Hz</i>	<i>Heave_L05hz</i>	<i>Heave_05_07hz</i>	<i>Heave_07_085hz</i>	<i>Heave_M085</i>	<i>Heave_L05hz</i>	<i>Heave_05_07hz</i>	<i>Heave_07_085hz</i>	<i>Heave_M085</i>
Above_Level	1	1	1	0	1	1	1	1
Below_Level	0	0	0	1	0	0	0	0

Table A14

Probability table of the node Acute Motion (MII) (Parent node: *Motion - Lateral Acceleration*).

Acute Motion (MII)		Probability		
<i>Motion - Lateral Acceleration</i>	<i>Lat_Acc_L012</i>		<i>Lat_Acc_012_025</i>	<i>Lat_Acc_M025</i>
Above_Level	0		0.67	1
Below_Level	1		0.33	0

Table A15

Probability table of the node Diffuse Vibration and Diffuse Vibration 2 (Parent node for the former: *Vibration*; parent node for the latter: *Vibration* 2).

Diffuse Vibration				Probability		
<i>Vibration</i>	<i>V_below_2hz</i>	<i>V_2_4hz</i>	<i>V_4_11hz</i>	<i>V_11_20hz</i>	<i>V_20_80hz</i>	<i>V_80_plus_hz</i>
Above_Level	0.01	0.01	0.01	0.01	0.01	0.01
Below_Level	0.99	0.99	0.99	0.99	0.99	0.99

Table A16

Probability table of the node Acute Vibration (Parent node: *Vibration*).

Acute Vibration				Probability		
<i>Vibration</i>	<i>V_below_2hz</i>	<i>V_2_4hz</i>	<i>V_4_11hz</i>	<i>V_11_20hz</i>	<i>V_20_80hz</i>	<i>V_80_plus_hz</i>
Above_Level	0	0.5	0.9	0.3	0.1	0
Below_Level	1	0.5	0.1	0.7	0.9	1

Table A17

Probability table of the node Diffuse Noise and Diffuse Noise 2 (Parent node for the former: *Noise*; parent node for the latter: *Noise* 2).

Diffuse Noise		Probability			
<i>Noise</i>	<i>s_L30dB</i>	<i>s_30_55dB</i>	<i>s_55_75dB</i>	<i>s_75_90dB</i>	<i>s_M90dB</i>
Above_Level	0.01	0.01	0.01	0.5	0.5
Below_Level	0.99	0.99	0.99	0.5	0.5

Table A18

Probability table of the node Acute Noise and Acute Noise 2 (Parent node for the former: *Noise*; parent node for the latter: *Noise 2*).

Acute Noise	Probability				
	<i>s_L30dB</i>	<i>s_30_55 dB</i>	<i>s_55_75 dB</i>	<i>s_75_90 dB</i>	<i>s_M90dB</i>
Noise					
Above_Level	0	0	0.5	0.8	1
Below_Level	1	1	0.5	0.2	0

Table A19

Probability table of the node GDF stressor and GDF stressor technical (Parent nodes for the former: *Diffuse Motion (MIS)*, *Diffuse Vibration*, *Diffuse Noise* and *Acute Noise*; parent nodes for the latter: *Diffuse Motion 2 (MIS_2)*, *Diffuse Vibration 2*, *Diffuse Noise 2* and *Acute Noise 2*).

GDF stressor	Probability							
	<i>Above_Level</i>				<i>Below_Level</i>			
<i>Diffuse Motion (MIS)</i>	<i>Above_Level</i>							
<i>Diffuse Vibration</i>	<i>Above_Level</i>				<i>Below_Level</i>			
<i>Diffuse Noise</i>	<i>Above_Level</i>		<i>Below_Level</i>		<i>Above_Level</i>		<i>Below_Level</i>	
<i>Acute Noise</i>	<i>Above_Level</i>	<i>Below_Level</i>	<i>Above_Level</i>	<i>Below_Level</i>	<i>Above_Level</i>	<i>Below_Level</i>	<i>Above_Level</i>	<i>Below_Level</i>
Active	1	1	1	1	1	1	1	1
Inactive	0	0	0	0	0	0	0	0

GDF stressor	Probability							
	<i>Above_Level</i>				<i>Below_Level</i>			
<i>Diffuse Motion (MIS)</i>	<i>Below_Level</i>							
<i>Diffuse Vibration</i>	<i>Above_Level</i>				<i>Below_Level</i>			
<i>Diffuse Noise</i>	<i>Above_Level</i>		<i>Below_Level</i>		<i>Above_Level</i>		<i>Below_Level</i>	
<i>Acute Noise</i>	<i>Above_Level</i>	<i>Below_Level</i>	<i>Above_Level</i>	<i>Below_Level</i>	<i>Above_Level</i>	<i>Below_Level</i>	<i>Above_Level</i>	<i>Below_Level</i>
Active	1	1	1	1	1	1	1	0
Inactive	0	0	0	0	0	0	0	1

Table A20

Probability table of the node GDF Physical effect (Parent nodes: *Acute Vibration* and *Acute Motion (MII)*).

GDF Physical effect	Probability			
	<i>Above_Level</i>		<i>Below_Level</i>	
<i>Acute Vibration</i>	<i>Above_Level</i>		<i>Below_Level</i>	
<i>Acute Motion (MII)</i>	<i>Above_Level</i>		<i>Below_Level</i>	
Yes	1	1	1	0
No	0	0	0	1

APPENDIX B. Quantification of the human behaviour in the presence of GDFs

(See [Tables B21–B28](#)).

Table B21

The conditional probability table of the node *C1-Detection, Assessment and Execution of simple action* (Parent nodes: *Helmsman Present*, *Attention Management Capability*, *GDF Physical Effect*).

C1-Detection, Assessment and Execution of simple action (DAAct)	Probability							
	Yes				No			
	Normal		Degraded		Normal		Degraded	
	Yes	No	Yes	No	Yes	No	Yes	No
Helmsman present								
Attention Management Capability								
GDF Physical Effect								
C1 performed	0.9999	0.9999	0.9999	0.9999	0.999	0.9995	0.9989	0.9994
C1 not performed	0.0001	0.0001	0.0001	0.0001	0.001	0.0005	0.0011	0.0006

Table B22

The conditional probability table of the node D1 – Verbal Communication on Safety Critical Data (Parent nodes: Attention Management Capability, C1-Detection, Assessment and Execution of simple action).

D1-Verbal Communication On Safety Critical Data	Probability			
	Normal (z1)		Degraded (z0)	
Attention Management Capability	<i>CI performed</i>		<i>CI not performed</i>	
C1-Detection, Assessment and Execution of simple action	<i>CI performed</i>		<i>CI not performed</i>	
D1 performed	0.994	0	0.9928	0
D1 not performed	0.006	1	0.0072	1

Table B23

The conditional probability table of the node Evasive Action (Parent nodes: Helmsman Present, C1-Detection, Assessment and Execution of simple action, D1 – Verbal Communication on Safety Critical Data).

Evasive Action	Probability							
	Yes				No			
Helmsman present	<i>CI performed (1)</i>				<i>CI not performed (0)</i>			
C1-Detection, Assessment and Execution of simple action	<i>Normal</i>				<i>Degraded</i>			
Attention Management Capability	<i>Normal</i>		<i>Degraded</i>		<i>Normal</i>		<i>Degraded</i>	
D1-Verbal Communication On Safety Critical Data	D1 Performed	D1 Not Performed	D1 Performed	D1 Not Performed	D1 Performed	D1 Not Performed	D1 Performed	D1 Not Performed
Evasive action executed	1	0.9	0.5	0.5	1	1	0	0
Evasive action not executed	0	0.1	0.5	0.5	0	0	1	1

Table B24

Probability table of the node Technical Failure (Parent node: Maintenance Task Performance).

Maintenance Task Performance (MTP)	Sufficient		Insufficient	
	No	1	0	0
Yes	0	0	1	1

Table B25

Probability table of the node Maintenance Task Performance (Parent node: Attention Management Capability 2).

Maintenance Task Performance	Probability	
	Normal (z1)	Degraded (z0)
Sufficient (m1)	0.9995	0.9994
Insufficient (m0)	0.0005	0.0006

Table B26

The conditional probability table of the node Collision (Parent nodes: Evasive Action of Another Ship, Evasive Action, Technical Failure).

Collision	Probability							
	Yes				No			
Evasive Action of Another Ship	<i>Executed</i>				<i>Not Executed</i>			
Evasive Action	<i>Executed</i>		<i>Not Executed</i>		<i>Executed</i>		<i>Not Executed</i>	
Technical Failure	No	Yes	No	Yes	No	Yes	No	Yes
Collision - yes	0	0	0	0	0	1	1	1
Collision - no	1	1	1	1	1	0	0	0

Table B27

Probability table of the node Grounding (Parent node: *Failure of another OOW, Evasive Action, Technical Failure*).

Grounding	Probability							
	Yes				No			
	Executed		Not Executed		Executed		Not Executed	
Technical Failure	No	Yes	No	Yes	No	Yes	No	Yes
Grounding - yes	0	1	1	1	0	1	0	1
Grounding - no	1	0	0	0	1	0	1	0

Table B28

A list of Error Producing Conditions (EPC) as defined in NARA, [51].

NARA No	EPC
1	Poor, ambiguous or ill-matched system feedback
2	Unfamiliarity, i.e. a potentially important situation which only occurs infrequently or is novel
3	A need to unlearn a technique and apply one which requires the application of an opposing philosophy
4	Time pressure
5	Low signal to noise ratio
6	A conflict between immediate and long-term objectives
7	No obvious means of reversing an unintended action, i.e. carrying out a complex task where the system is unforgiving of errors
8	A means of suppressing or over-riding information or features which is too easily accessible
9	Operator inexperience
10	Cognitive overload, particularly one caused by simultaneous presentation of non-redundant information
11	No obvious way of keeping track of progress during an activity
12	Shortfalls in the quality of information conveyed by procedures
13	Difficulties caused by poor shift hand-over practices and/or team co-ordination problems or friction between team members
14	An incentive to use other more dangerous procedures to achieve long-term objectives
15	Poor environment
16	High emotional stress and effect of ill health
17	Low workforce morale or adverse organizational environment
18	Operator under-load/boredom

References

[1] Mazaheri A, Montewka J, Kujala P. Modeling the risk of ship grounding – a literature review from a risk management perspective. *WMU J Marit Aff* 2014;13(2):269–97.

[2] Papanikolaou A, Bitha K, Eliopoulou E, and Ventikos NP. Statistical analysis of ship accidents occurred in the period 1990-2012 and assessment of safety level of ship types; 2014.

[3] Chauvin C, Lardjane S, Morel G, Clostermann J-P, Langard B. Human and organisational factors in maritime accidents: analysis of collisions at sea using the HFACS. *Accid Anal Prev* 2013;59:26–37.

[4] Mullai A, Paulsson U. A grounded theory model for analysis of marine accidents. *Accid Anal Prev* 2011;43(4):1590–603.

[5] Leva MC, Friis-Hansen P, Ravn ES, Lepsoe A. SAFEDOR: A practical approach to model the action of an officer of the watch in collision scenarios. In: *European safety and reliability conference; Safety and reliability for managing risk, ESREL ; 2006*. pp. 2795–2804.

[6] Martins MR, Maturana MC. Application of Bayesian Belief networks to the human reliability analysis of an oil tanker operation focusing on collision accidents. *Reliab Eng Syst Saf* 2013;110:89–109.

[7] Kim MC, Seong PH, Hollnagel E. A probabilistic approach for determining the control mode in CREAM. *Reliab Eng Syst Saf* 2006;91(2):191–9, [Feb.].

[8] Sotiralis P, Ventikos NP, Hamann R, Golyshev P, Teixeira AP. Incorporation of human factors into ship collision risk models focusing on human centred design aspects. *Reliab Eng Syst Saf* 2016.

[9] May M. Cognitive aspects of interface design and human-centered automation on the ship bridge: the example of ARPA/ECDIS integration. In: *Proceedings of the international conference on people in control (Human Interfaces in Control Rooms, Cockpits and Command Centres); 1999*, vol. 1999. p. 394–399.

[10] Sauer J, Wastell DG, Hockey GRJ, Crawshaw CM, Ishak M, Downing JC. Effects of display design on performance in a simulated ship navigation environment. *Ergonomics* 2002;45(5):329–47.

[11] *Handbook of control room design and ergonomics: a perspective for the future, Second Edition*. CRC Press, 2008.

[12] Papanikolaou A, editor. *Risk-based ship design: methods, tools and applications*. Berlin, Heidelberg: Springer-Verlag; 2009.

[13] DNV, *Formal Safety Assessment – Large Passenger Ships, ANNEX II: Risk Assessment – Large Passenger Ships – Navigation*; 2003.

[14] Breinholt C, Ehrke K-C, Papanikolaou A, Sames PC, Skjong R, Strang T, Vassalos D, Witolla T. SAFEDOR–The implementation of risk-based ship design and approval. *Procedia – Soc Behav Sci* 2012;48:753–64.

[15] Vassalos D. Design for Safety: risk-based design life-cycle risk management. [Online]. Available: ([http://www.rina.org.uk/hres/Design for Safety Risk-Based Design Life-Cycle Risk Management.pdf](http://www.rina.org.uk/hres/Design%20for%20Safety%20Risk-Based%20Design%20Life-Cycle%20Risk%20Management.pdf)). [accessed 11.01.16].

[16] Konovessis D. An investigation on cost-effective tanker design configurations for reduced oil outflow. *Ocean Eng* 2012;49:16–24.

[17] Guarín L, Konovessis D, Vassalos D. Safety level of damaged RoPax ships: risk modelling and cost-effectiveness analysis. *Ocean Eng* 2009;36(12–13):941–51.

[18] Guarín L, Logan J. Risk fire screening of passenger ships layout in support of the assessment of alternative design and arrangements. In: *Proceedings of the design and operation of passenger ships*; 2011.

[19] Spanos D, Papanikolaou A. On the time dependent survivability of ROPAX Ships. In: *Proceedings of the 11th International Ship Stability Workshop*, ; 2010. pp. 143–147.

[20] Papanikolaou A, Eliopoulou E. On the development of the new harmonised damage stability regulations for dry cargo and passenger ships. *Reliab Eng Syst Saf* 2008;93:1305–16.

[21] Mallam SC, Lundh M, MacKinnon SN. Integrating human factors & ergonomics in large-scale engineering projects: Investigating a practical approach for ship design. *Int J Ind Ergon* 2015;50:62–72.

[22] Chauvin C, Le Bouar G, Renault C. Integration of the human factor into the design and construction of fishing vessels. *Cogn Technol Work* 2007;10(1):69–77.

[23] Rumawas V. *Human factors in ship design and operation: experimental learning*, Norwegian University of Science and Technology; 2015.

[24] Fireman S, Lossa A, Singer D. Fuzzy logic single-objective optimization using multiple criteria to understand human factors on early stage design. In: *Proceedings of*

- the 15th international conference on computer and IT applications in the maritime industries; 2016. pp. 481–490
- [25] Dobie TG. Critical significance of human factors in ship design. In: Proceedings of the 2003 RVOC meeting; 2003.
- [26] Kivimaa S, Rantanen A, Nyman T, Owen D, Garner T, Davies B. Ship motions, vibration and noise influence on crew performance and well-being studies in FAROS project. In: Transport research arena conference TRA; 2014.
- [27] Montewka J. Summarizing literature review (Human Factors in Risk-Based Ship Design Methodology, Project no 314817), Espoo, 2014.
- [28] Rantanen A. Quantitative models of crew performance linked to ship motions. Deliverable 3.2. Human factors in risk-based ship design methodology project no. 314817, Espoo, 2013.
- [29] Hancock PA. A dynamic model of stress and sustained attention. *Hum Factors* 1989;31(5):519–37.
- [30] Robert G, Hockey J. Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. *Biol Psychol* 1997;45(1):73–93.
- [31] Young MS, Stanton NA. Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Hum Factors* 2002;44(3):365–75.
- [32] Darwiche A. Modeling and reasoning with bayesian networks, 1st ed.. Cambridge University Press; 2009.
- [33] Montewka J, Ehlers S, Goerlandt F, Hinz T, Tabri K, Kujala P. A framework for risk assessment for maritime transportation systems – a case study for open sea collisions involving RoPax vessels. *Reliab Eng Syst Saf* 2014;124:142–57.
- [34] Montewka J, Goerlandt F, Kujala P. On a systematic perspective on risk for formal safety assessment (FSA). *Reliab Eng Syst Saf* 2014;127:77–85.
- [35] Zhang D, Yan XP, Yang ZL, Wall A, Wang J. Incorporation of formal safety assessment and Bayesian network in navigational risk estimation of the Yangtze River. *Reliab Eng Syst Saf* 2013;118:93–105.
- [36] Wang C, Mosleh A. Qualitative-quantitative bayesian belief networks for reliability and risk assessment, 2010, p. 1–5.
- [37] Li KX, Yin J, Bang HS, Yang Z, Wang J. Bayesian network with quantitative input for maritime risk analysis. *Transportmetrica* 2012;1–30.
- [38] Pitchforth J, Mengersen K. A proposed validation framework for expert elicited Bayesian networks. *Expert Syst Appl* 2013;40(1):162–7.
- [39] Goerlandt F, Khakzad N, Reniers G. Validity and validation of safety-related quantitative risk analysis: a review. *Saf Sci Spec Issue Risk Anal Valid Trust Risk Manag* 2016.
- [40] IMO. Revised guidelines for formal safety assessment (FSA) for use in the IMO rule-making process. MSC-MEPC.2/Circ.12, LONDON, UK, 2013.
- [41] Jakeman AJ, Letcher RA, Norton JP. Ten iterative steps in development and evaluation of environmental models. *Environ Model Softw* 2006;21(5):602–14.
- [42] Blackman HS, Gertman DI, Boring RL. Human error quantification using performance shaping factors in the SPAR-H method. In: Proceedings of the human factors and ergonomics society annual meeting, 2008, vol. 52, no. 21, pp. 1733–1737.
- [43] Montewka J, Goerlandt F, Innes-Jones G, Owen D, Hifi Y, Porthin M. Quantifying the effect of noise, vibration and motion on human performance in ship collision and grounding risk assessment. In: PSAM 2014 – Probabilistic safety assessment and management; 2014.
- [44] Bos JE. How motions make people sick such that they perform less: a model based approach. In: Proceedings of the NATO RTO/AVT-110 symposium on habitability of combat and transport vehicles: noise, vibration and motion; 2004. pp. 1–11
- [45] Colwell JL. Modeling ship motion effects on human performance for real time Simulation. *Nav Eng J* 2005;117(1):77–90.
- [46] Jones DM, Macken WJ, Mosdell NA. The role of habituation in the disruption of recall performance by irrelevant sound. *Br J Psychol* 1997;88(4):549–64.
- [47] Crossland P, Rich KJN. A method for deriving MII criteria. In: Human factors in ship design and operation; RINA international conference; 2000.
- [48] Conway GE, Szalma JL, Hancock PA. A quantitative meta-analytic examination of whole-body vibration effects on human performance. *Ergonomics* 2007;50(2):228–45.
- [49] Hollnagel E. Cognitive reliability and error analysis method (CREAM). Elsevier Science; 1998.
- [50] He X, Wang Y, Shen Z, Huang X. A simplified CREAM prospective quantification process and its application. *Reliab Eng Syst Saf* 2008;93(2):298–306.
- [51] CorporateRisk Associates. A user manual for the nuclear action reliability assessment (NARA) human error quantification technique. Report No 2, Surrey, UK; 2011.
- [52] Bowditch N. The American practical navigator, 2015th ed.. Bethesda, Maryland, USA: National Imagery and Mapping Agency; 2002.
- [53] International Chamber of Shipping . Bridge procedures guide. London: Marisee Publications; 2007.
- [54] Rowe WD. Understanding uncertainty. *Risk Anal* 1994;14(5):743–50.
- [55] Winkler RL. Uncertainty in probabilistic risk assessment. *Reliab Eng Syst Saf* 1996;54(2–3):127–32.
- [56] Coupé VMH, van der Gaag LC. Properties of sensitivity analysis of Bayesian belief networks. *Ann Math Artif Intell* 2002;36:323–56.
- [57] Flage R, Aven T. Expressing and communicating uncertainty in relation to quantitative risk analysis. *Reliab Risk Anal Theory Appl* 2009;2(13):9–18.
- [58] Montewka J, Goerlandt F, Innes-Jones G, Owen D, Hifi Y, Porthin M. Quantifying the effect of noise, vibration and motion on human performance in ship collision and grounding risk assessment. In: Proceedings of the 12th probabilistic safety assessment and management conference, PSAM ; 2014.
- [59] Goerlandt F, Reniers G. On the assessment of uncertainty in risk diagrams. *Saf Sci* 2016;84:67–77.
- [60] Drouin M, Parry G, Lehner J, Martinez-Guridi G, LaChnace J, Wheeler T. Guidance on the treatment of uncertainties associated with PRAs in risk-informed decision making (NUREG-1855), Upton, New York; 2009.
- [61] Paté-Cornell ME. Uncertainties in risk analysis: six levels of treatment. *Reliab Eng Syst Saf* 1996;54(2–3):95–111.
- [62] Abdellaoui M, Luce RD, Machina M, Munier B. Uncertainty and risk - mental, formal, experimental. Heidelberg, Berlin: Springer; 2007.
- [63] Dutta P. Uncertainty modeling in risk assessment based on dempster–shafer theory of evidence with generalized fuzzy focal elements. *Fuzzy Inf Eng* 2015;7(1):15–30.
- [64] Aven T. Quantitative risk assessment: the scientific platform. Cambridge University Press; 2011.
- [65] Merrick JRW, Van Dorp JR, Dinesh V. Assessing uncertainty in simulation-based maritime risk assessment. *Risk Anal* 2005;25(3):731–43.
- [66] Pedroni N, Zio E, Ferrario E, Pisanisi A, Couplet M. Propagation of aleatory and epistemic uncertainties in the model for the design of a flood protection dike, PSAM 11 & ESREL; 2012. 25-Jun-2012. p. 1–10.
- [67] Sormunen O-VE, Goerlandt F, Häkkinen J, Posti A, Hänninen M, Montewka J, Ståhlberg K, Kujala P. Uncertainty in maritime risk analysis: extended case study on chemical tanker collisions. *Proc Inst Mech Eng Part M: J Eng Marit Environ* 2015;229(3):303–20.
- [68] Flage R, Aven T, Zio E, Baraldi P. Concerns, challenges, and directions of development for the issue of representing uncertainty in risk assessment. *Risk Anal* 2014;34(7):1196–207.
- [69] Aven T, Zio E, Baraldi P, Flage R. Uncertainty in risk assessment: the representation and treatment of uncertainties by probabilistic and non-probabilistic methods. Wiley; 2014.
- [70] O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, Oakley JE, Rakow T. Uncertain judgments: eliciting experts' probabilities. Wiley; 2006.
- [71] Hänninen M, Mazaheri A, Kujala P, Montewka J, Laaksonen P, Salmiovirta M, Klang M. Expert elicitation of a navigation service implementation effects on ship groundings and collisions in the Gulf of Finland. In: Proceedings Inst. Mech. Eng. Part O: J. Risk Reliab., no. 228; 2014. p. 19–28.
- [72] Fowler TG, Sorgård E. Modeling ship transportation risk. *Risk Anal* 2000;20(2):225–44.
- [73] Brandowski A. Estimation of the probability of propulsion loss by a seagoing ship based on expert opinions. *Pol Marit Res* 2009;1(59):73–7.
- [74] Hänninen M, Kujala P. Influences of variables on ship collision probability in a Bayesian belief network model. *Reliab Eng Syst Saf* 2012;102(0):27–40.
- [75] Anderson A, Greal M, Thomson J, Butler S, Benedict K, Linnenbecker M, Tuschling G. Results of physical experiments. Deliverable 7.2, 2015.
- [76] Leva MC, Leva MC. Model of the Operator. SAFEDOR, deliverable for Subtask 2.4. 2.; 2006.
- [77] Martins MR, Maturana MC. Human error contribution in collision and grounding of oil tankers. *Risk Anal* 2010;30(4):674–98.
- [78] Asami M, Kaneko F. Development of vessel collision model based on naturalistic decision making model. In: Proceedings of the Collision and Grounding of Ships and Offshore Structures; 2013.
- [79] Carson ER, Flood RL. Model validation: philosophy, methodology and examples. *Trans Inst Meas Control* 1990;12(4):178–85, [Jan.].
- [80] R. I. S. M , Valentina Di Pasquale SR. Operations Management. InTech; 2013.