Effects of the Background and Experience on the Experts' Judgments through Knowledge Extraction from Accident Reports

Noora Hyttinen[a], Arsham Mazaheri[b*], and Pentti Kujala[c]
[a] Aalto University, Department of Applied Mathematics, School of Science, Espoo, Finland
[b] Aalto University, Department of Applied Mechanics, School of Engineering, Espoo, Finland and Kotka Maritime Research Center (Merikotka), Kotka, Finland
[c] Aalto University, Department of Applied Mechanics, School of Engineering, Espoo, Finland

Abstract: Available risk models for maritime risk analysis are not proper enough for risk management purposes as they are not evidence-based. One of the sources of evidence that can be used for accident modeling is the accident reports. The reports need to be reviewed to extract the presented knowledge. This study investigates how the differences in the background and expertise of the reviewers can affect the extracted knowledge from the accident reports. The study is conducted by utilizing three-round Delphi method and using two test groups as researchers and mariners to review four grounding accident reports prepared by Finnish Accident Investigation Board. The results of the study show that although neither of the groups have superiority over the other with regard to the extracted knowledge, there are some categories that are chosen more frequently by specific group. Mariners chose more often the causes related to navigation and the actions of the crew, while the researchers tend to see more organizational and environmental related causes. Thus, the background of the reviewer should be considered in evidence-based modeling, as it affects the resulting models and thus the implementing risk control options suggested by the constructed models.

Keywords: Grounding, Accident Reports, Knowledge Extraction, Belief Networks, Hamming-Distance

1. INTRODUCTION

Marine accidents are not only seen as the threat to the maritime transportation itself because of the posing risks to the human lives and the environment, they are also threats to the international trade and industry by interrupting the supply networks [1]. Thus, it is of great importance to keep the likelihood and consequence of an accident as low as reasonably possible. In order to do so, risk managers usually use risk models to understand the system and to mitigate the involved risk by implementing proper risk control options. In this regard, a suitable model for risk management purposes should reflect the true available knowledge on the system with satisfying accuracy [2]. However, currently available risk models for maritime risk analysis are not proper enough for risk management purposes as they are not based on the evidences rather on the intuition of the developers, thus not presenting the true available knowledge of the system [3]. To mitigate this flaw, evidence-based modeling that deal with real accident scenarios, as opposed to imaginary scenarios, is encouraged [3,4]. One of the rich sources of the evidence useful for modeling is the accident reports prepared by the accident investigation boards. When building a database required for modeling, the required causal networks of events need to be extracted from the accident reports. Often researchers, as system experts, read the accident reports or use an algorithm [5-7] to identify the causal networks of the accidents. The problem with researchers reading the reports is that they may not necessarily have the appropriate expertise to get the needed knowledge from the reports. There is also a possibility of misunderstanding because the reports are written in natural language [8]. When reading the accident reports it is likely that different people perceive the events of the accident differently [9].

* Corresponding author: arsham.mazaheri@aalto.fi

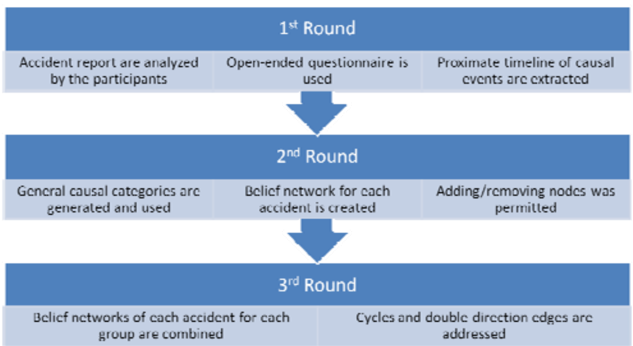
To reduce the uncertainty in mathematical models that are built based on the reports, it is necessary to study if certain individuals are better equipped to extract the knowledge presented in the accident reports. Thus, since the expertise most likely gives the expert better understanding of the events leading to an accident, it seems necessary to know how the background and expertise of the reader affects the way she sees the causal network of events and thus affects the extracted knowledge. The general opinion may be that the people with insight in the field of the researched subject are more qualified to review accident reports to extract the knowledge about the accident causes. The mariners may be less aware of accident theories and accident causality whereas the researchers may have less understanding of the ship's operational factors and navigation. Researchers that study accident theories and the causes of accidents may see connections between the causes better and are better equipped to review large amounts of data; while mariners may find the individual causes from the accident reports better. Therefore, the objective of this study is to find out how a person's background and knowledge affect the causal network that is extracted from accident reports and the mathematical model that is built based on that– especially whether researchers can identify the same causal networks as mariners with years of nautical experience.

The outline of this paper is as follow: the research procedure, the methods, and the reports used for the study together with the composition of the participants are presented in the next section; the results of the study based on the collected responses from the participants are summarized in Section 3; the results are then analyzed and discussed in Section 4; and finally the paper is concluded in Section 5.

2. METHODOLOGY

The study focuses on grounding accidents. Delphi method [10] with three rounds was implemented to perform the study and to collect the responds of the participants (**Figure 1**). The first round consisted of a questionnaire with open-ended questions. Open-ended questions were used in order not to influence the participants' judgments. The aim of the first part was to get the participants to extract proximate causal events from accident reports. Thus, four grounding accident reports prepared by the Finnish Accident Investigation Board (**Table 1**) were chosen from among the reports that have English translations and sent to the participants for the analysis. The participants were asked to review the accident reports and collect the causes of the accident and list them in a timeline manner that ends to a grounding accident.

Figure 1: Three-round Delphi method is used in the study



The participants were two groups of experts. The first group consisted of six marine specialists with nautical education and work experience in either navigation or marine accident investigation with average of 16 years of nautical experience. The second group consisted of six researchers in the maritime transportation domain with the average of 3.5 years of research experience but no formal nautical education or experience. The questionnaire of the first round was pilot tested by a person who had both nautical education and research work experience.

After all answers of the first round were collected, the extracted causes are grouped into 22 general categories (**Table 2**). Categories from Ref. [11] and Ref. [12] are used as guidelines in building the categories related to human factors. The three largest human factors in marine accidents are fatigue, inadequate communications and cooperation between pilot and bridge crew, and inadequate technical knowledge [11]. All of these three causes were found in the studied accident reports. The above causes can be classified as human factors and they are mostly errors that are made by the bridge crew. In addition there were some organizational factors in the studied accident reports. These are mostly errors made by the shipping companies and other outside operators. In addition to human and organizational factors, environmental factors were thought to have caused groundings. Environmental factors are considered to be factors that cannot be controlled by the crew or the shipping company.

Table 1: Grounding accident reports used in the study

Name of the vessel	Time of the accident	Location
MV EMSRunner	11.12.2009	Kalajoki
MS Pauline Russ	20.01.2005	Hanko
MS Claudia	23.10.2007	Tornio
MS Superfast VII	12.11.2004	Hanko

Table 2: List of the categories that were found in the reviewed accident reports

Categories	Short Description of the causes in the category
Human Factors	
Fatigue	Fatigue or other personal reasons, workload
Faulty practices	Faulty standards, policies or practices
Hazardous natural environment	Not adjusting operations based on hazardous natural environment (sea clutter, wind)
Inadequate communications	Lacking communications and cooperation, route plan not discussed
Inadequate route plan	Route plan missing, not on the radar, inadequate, bypassing distances missing
Inappropriate use of navigational equipment	Lack of or incorrect inappropriate use of navigational equipment
Incompetence	Lack of competence, knowledge and ability, other violation of good seamanship practices
Lack of redundancy	Lack of redundancy
Lack of situational awareness	Lack of situational awareness or monitoring
Manning	Too few people on bridge
Mistake	Complacency, mistake made by crew, decisions based on inadequate information
Route plan not followed	Route plan not followed
Wrong steering decision	Wrong steering decision, corrective manoeuvres were made too late, the incorrect use of tugboats
Organizational Factors	
Lack of guidelines	Lack of guidelines from the shipping company, including guidelines on using tugboats, wind limitations etc. (SMS), lack of risk analysis, systems improving ship operations, feedback
Lack of training	Lack of simulator practice/other education/inadequate instructions for using the navigation equipment
Lack of VTS	Lack of help from VTS or other outside operator, VTS not involved
Poor design of automation	Poor design of automation, rudimentary or inadequate navigation or other equipment
Schedule	Schedule, time pressure from shipping company, productivity
Environmental Factors	
Bad visibility	Darkness, fog, dazzling lights of another vessel, sun glare
Fairway	Wrongly marked shallow, fairway or narrow fairway, pilot station placed unfavourably
Outside distractions	Outside factors, dredging or other disturbing factor on fairway
Wind	Wind

At the second round, the created categories of causal events were sent back to the participants and they have been asked to structure a belief network (BN) for each accident using the general categories as nodes, in the way that reflects their beliefs on the causal network that ends to the accident. A brief instruction of BN was attached to the questionnaire to help those who were not familiar with the concept. To make this round easier, each participant has her own questionnaire with categories (nodes) that were related to the causal events that she has reported in the first round. A list that includes all categories from all participants and all four studied accident reports was also attached to the questionnaire. The list gives some directions on how other participants interpreted the causes in the accident reports. All of the categories were in the same list to prevent the participants of seeing which causes were thought to have been present in which accident. Adding and removing nodes from the list was permitted in the process of constructing BN, though only few participants used this possibility. In general, there were no limitations in the instructions on how many nodes there should be in one network; therefore the number of nodes varied from one (excluding the node “grounding”) to eleven nodes per BN. As the result, none of the structured BNs were exactly the same for same accident and there were some BNs of the same accident that did not have even a single similar node in them.

After collecting BNs from the two groups for each accident, the networks for each accident were combined for each group to make two networks for each accident, one for mariners and one for researchers. The network of a group was built from the majority opinion; meaning if a node or an edge (arc) is in at least half of the participants’ networks it is selected to the group’s network. In case of a tie, the participants who have more experience in accident investigation or accident research are weighted 50% higher than the other participants. The resulted grouped networks were then used for comparing the two groups.

Since combining networks may create cycles and double direction edges, the third round of the questionnaire is used to eliminate those problems in the combined networks. Thus, the participants were asked about the contradicting edges in their own groups’ combined networks. The third questionnaire consisted of multiple choice questions about the directions of the edges between certain nodes. The assumption in the third part was that the nodes in question were causes of the accident. The participants were asked to either choose the direction of the edge or not to use an edge.

3. RESULTS

The structured BNs are compared in three ways: 1) the distances between different individual BNs are calculated using a redefined version of the Hamming-Distance. The distances are compared in order to find the general existing differences in the views of each group; 2) the nodes of the individual and combined BNs are compared by calculating the appearance frequency of the nodes in the BNs of the two test groups. This helps us to find the most common causes that are detected in the reports by each group, and to understand the possible existing differences between their views; 3) the edges of the individual and combined BNs are compared by assessing the nodes that are connected with each edge. This helps us to find believes of each group on the dependencies of different causes, and to understand the possible existing differences between their views.

3.1. Distances

The networks of the two groups were compared with each other by calculating an alternative distance to the Hamming-Distance. The Hamming-Distance describes the number of changes that have to be made to a network for it to turn into the one that is being compared with [13]. In calculating the Hamming-Distance, both networks have to have the same nodes. Moreover, a reference network, which would be the network that other networks are compared to, is needed. The problem in this study was that the reference networks did not exist. Therefore, instead of using the original Hamming-Distances (***Eq.1***), a different method of calculating differences of two networks, as Alternative-Distance, was introduced (***Eq.4***) and used. Hamming-Distance is calculated as:

$$H = A + D + I \tag{1}$$

where A is the added, D deleted, and I incorrectly oriented edges in the constructed network in comparison to the reference network [14]. The added edges in the Hamming-Distance represent the edges that need to be removed from the BN that is being compared with the reference network. The deleted edges represent the edges that the reference network has but the compared BN does not.

The research setting of this study was such that the created networks did not include all of the possible nodes. The nodes had as much significance in this study as the edges. Neither of the BNs is considered to be the reference network. Thus, if we rename the reference network as BN1 and the BN that it is compared with as BN2, we can say that the added edges are extra edges in BN2 and the deleted edges are extra edges in BN1. The changes in the names of the variables are shown in **Table 3**.

Table 3: Changes in the names of the variables from the Hamming-Distance, when using the introduced Alternative-Distance

Hamming-Distance (*H*)	**Alternative-Distance (*Dist*)**
Reference network	BN_1
Compared network	BN_2
Added edges (A)	Extra edges in BN_2 (E_{BN2})
Deleted edges (D)	Extra edges in BN_1 (E_{BN1})
Incorrectly oriented edges (I)	Differently oriented edges (I)

Because the BNs in the study are of different sizes, the extra edges are divided so that they are proportional to the number of all possible edges in the BN. The edges that are left out of the BNs also tell us about the dependencies of the variables, and they should also be taken into account in calculating the distance between two BNs.

The Hamming-Distance is divided by all possible edges in a BN (***Eq.2***) that includes all nodes from the BNs that are compared with each other (see ***Eq.3***). The node "*grounding*" is not counted to the total number of nodes in the network as it is included in every network. Because of that, the number of all possible places for edges in the combined BN can be calculated as:

$$Edg = \sum_{i=1}^{n} i \tag{2}$$

where n is the number of different nodes in the two BNs. *Edg* tells us how many edges are possible in the combined BN at the same time. It is possible to have only one edge between two nodes, otherwise a cycle would be created. It is also possible to have edges between all nodes without having any cycles in the network. Therefore, the Hamming-Distance for our study should be calculated as:

$$Dist = \frac{E_{BN1} + E_{BN2} + I}{Edg} \tag{3}$$

In this study we were also interested in how many nodes the BNs have in common, because the BNs that have fewer different nodes can be considered to be more similar than those that have more different nodes. Therefore, to account the differences in the nodes of the BNs, we have added an extra part to ***Eq. 3***, and thus have introduced the *Alternative-Distance (**Eq.4**)*.

$$Dist = \frac{E_{BN1} + E_{BN2} + I}{Edg} + \frac{N_{BN_1} + N_{BN_2}}{Nod} \tag{4}$$

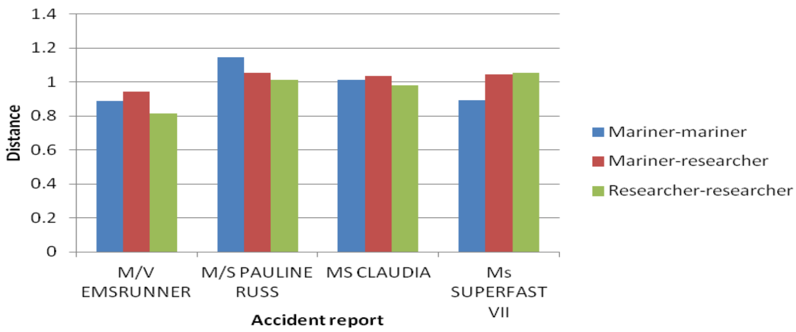
where N_k is extra nodes in network k and *Nod* is the number of nodes in the combined network.

The difference of Alternative-Distance to the definition of the Hamming-Distance is that ***Eq.4*** calculates the part of nodes and edges that have to be removed (extra edges and nodes) or changed (differently oriented edges) to form identical networks, rather than only the total amount of edges that need to be removed (added edges), added (deleted edges) or changed (incorrectly oriented edges) as in the Hamming-Distance.

The minimum value given by ***Eq.4*** when the BNs are identical is zero. When none of the edges are same, that is $E_{BN1} + E_{BN2} = Edg$ and $I = 0$, the maximum of the first part of ***Eq.4*** is one. For all edges to be different all nodes have to be the same. The maximum value of the second part of ***Eq.4*** is also one when none of the nodes are same, that is $N_{BN1} + N_{BN2} = Nod$. Thus the Alternative-Distance between two BNs cannot have the value of more than two, though it can never be exactly two either.

Figure 2 shows the average Alternative-Distances between individual BNs of the participants. The mariner-mariner and researcher-researcher distances were calculated by comparing each participant's BN to one another inside each group and calculating the average of all those distances. The mariner-researcher distance was calculated similarly by comparing every mariner's BN with each researcher's BN. The average distances show the variation inside each group and the variation between the groups. The lower average distance indicates that when the individual BNs in the group are compared with each other they have more similarities than in the groups that have higher average distances. **Figure 2** shows that the researchers' BNs are more uniform than the mariners'; meaning less variation between the individual researchers is seen than between individual mariners. This may be the sign of more consistency in the common knowledge of the researchers than the mariners, which can be the result of the smaller differences in the years of experience of the individuals in the researchers' group that is five years in compare with the mariners' that is 17 years. Nevertheless, in general, no specific trends or significant differences between mariners and researchers can be recognized from **Figure 2**, which indicates that neither of the groups have superiority over the other.

Figure 2: Distances on average between individual BNs



3.2. Nodes

The nodes of the BNs were collected in **Table 4**. The table shows the number of participants from each test group who detected the cause as an affecting factor in the reviewed accident reports. The maximum of each group is six. The differences between the groups can be better compared with the combined results of all reports and their differences in the last three columns on the right (**Table 4**).

In order to find which categories more frequently are chosen by each group as well as in general, the frequencies of each category are calculated using ***Eq.5*** and are shown in **Table 5.**

$$Frequency = \frac{Times\ that\ the\ cause\ occured\ in\ all\ four\ reports'BNs}{4 * Number\ of\ participants\ in\ the\ group} \tag{5}$$

The rank order of the frequencies of the categories for the mariners and the researchers in **Table 5** shows Spearman rank correlation of 0.71. This shows that the order of the causes almost matches between the groups. There are some clear differences though, as is seen from **Table 4**. For instance, *inadequate route plan* was thought to cause groundings three times more often by mariners than researchers. This is the biggest difference between the groups. Other categories that the mariners had chosen more frequently in their BNs are *wrong steering decision*, *poor design of automation*, and *lack of training*. **Table 4** also shows that, for instance, *incompetence* was chosen only by the mariners;

while *mistake*, *lack of guidelines, wind,* and *bad visibility* are chosen more frequently by the researchers than the mariners as a cause of accident. In general, it can be seen that the causes that are more related to navigation and ship handling as well as actions of the crew are detected and chosen more frequently by the mariners, while the researchers tend to detect and choose organizational and environmental causes more frequently.

Table 4: The frequency of each category in all accident reports, recognized by M = mariners and R = researchers

		MV EMSRUNNER		MS PAULINE RUSS		MS CLAUDIA		MS SUPERFAST VII		Total		Difference (R – M)
	Category	**M**	**R**	**M**	**R**	**M**	**R**	**M**	**R**	**M**	**R**	
More frequently chosen by the MARINERS	Inadequate route plan	6	0	3	0	4	4	0	0	13	4	-9
	Wrong steering decision	2	0	3	4	0	0	5	0	10	4	-6
	Lack of situational awareness	4	2	1	0	2	0	2	2	9	4	-5
	Poor design of automation	4	1	0	0	3	1	0	0	7	2	-5
	Incompetence	2	0	1	0	1	0	1	0	5	0	-5
	Lack of training	0	0	0	0	0	0	4	1	4	1	-3
	Manning	0	0	3	1	0	0	1	0	4	1	-3
Almost INDIFFERENCE between mariners and researchers	Lack of VTS	6	4	0	0	0	2	1	0	7	6	-1
	Outside distractions	0	0	0	0	3	2	0	0	3	2	-1
	Route plan not followed	1	0	0	0	0	0	0	0	1	0	-1
	Faulty practices	3	5	4	3	0	1	5	3	12	12	0
	Inappropriate use of navigational equipment	1	2	0	0	5	3	4	5	10	10	0
	Hazardous natural environment	0	0	3	4	2	1	3	3	8	8	0
	Fairway	4	5	0	0	0	1	1	0	5	6	1
	Lack of redundancy	0	2	0	0	1	0	0	0	1	2	1
More frequently chosen by the RESEARCHERS	Schedule	2	2	0	0	0	2	0	0	2	4	2
	Fatigue	1	3	0	0	0	0	0	0	1	3	2
	Inadequate communications	3	6	4	5	5	3	2	3	14	17	3
	Wind	0	0	1	2	0	0	1	3	2	5	3
	Bad visibility	1	2	0	0	3	6	0	0	4	8	4
	Lack of guidelines	0	1	3	3	2	5	5	6	10	15	5
	Mistake	3	6	2	3	0	1	0	0	5	10	5

There are some drawbacks of such generalization over the whole groups though, which one may want to consider. For example, none of the researchers thought that incompetence was a cause in the studied accidents, while only two of the six mariners thought that incompetence was a cause of the groundings; one of them chose incompetence to all of their BNs and the other to one BN. This may represent the opinion of only one mariner rather than the opinion of the whole group. For some other categories that are chosen too seldom, like *route plan not followed* or *lack of redundancy*, the drawn conclusions might be even weaker especially if they are divided between more than one accident. The differences are as likely caused by randomness than the differences in the knowledge of the test groups.

Table 5: Frequencies [0-1] of each category from all four accident reports

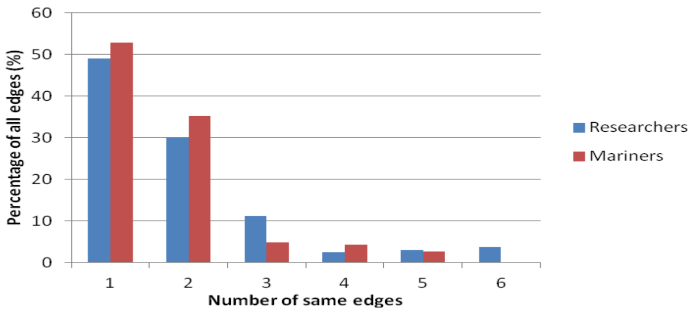
Category	Mariners	Researchers	Total
Inadequate communications	0,583	0,708	0,646
Lack of guidelines	0,417	0,625	0,521
Faulty practices	0,500	0,500	0,500
Inappropriate use of navigational equipment	0,417	0,417	0,417
Inadequate route plan	0,542	0,167	0,354
Hazardous natural environment	0,333	0,333	0,333
Mistake	0,208	0,417	0,313
Wrong steering decision	0,417	0,167	0,292
Lack of situational awareness	0,375	0,167	0,271
Lack of VTS	0,292	0,250	0,271
Bad visibility	0,167	0,333	0,250
Fairway	0,208	0,250	0,229
Poor design of automation	0,292	0,083	0,188
Wind	0,083	0,208	0,146
Schedule	0,083	0,167	0,125
Incompetence	0,208	0,000	0,104
Manning	0,167	0,042	0,104
Outside distractions	0,125	0,083	0,104
Lack of training	0,167	0,042	0,104
Fatigue	0,042	0,125	0,083
Lack of redundancy	0,042	0,083	0,063
Route plan not followed	0,042	0,000	0,021

3.3. Edges

The comparison of the edges is more complicated than the nodes as there are much more combinations. The edges in the BNs also vary generally more than the nodes because all of the participants do not have all same nodes in their BNs. Assuming that identifying the causes from accident reports does not require much knowledge on the interdependencies of the causes, the driven results from the nodes are considered to be more reliable than what can be driven from the edges. This is because constructing a BN by adding directed edges to the causes may need prior knowledge on causality, BN, and accident theories.

With that being said, looking at the most frequent edges from each node we can compare the structures of the BNs. We concentrated on the top 12 most frequent nodes in **Table 5**, where both test groups have in at least four of their individual BNs. The less frequent nodes have mostly single edges connected to them. **Figure 3** shows how many participants in each test group on average chose same edges to their BN of the same accident. It can be seen that researchers are more unanimous about the edges in their networks than the mariners, which again may be the result of the differences in the years of experience of the individuals in each test group.

Figure 3: Number of same edges in one report in relation to all edges



The dependencies in the BNs can also be viewed in terms of how often there is an edge between certain pair of nodes (in both directions) when both nodes are in the same BN. **Table 6** shows the dependencies between the most frequent nodes from **Table 5**. The numbers in **Table 6** show the fractions of the number of the BNs in a group that have specific pair of nodes and have indicated the dependency between that pair of nodes. For instance, there is dependency between fairway and lack of situational awareness in all of the researchers' BNs (100%), but in only 80 % of the mariners' BNs that have both of the two nodes (see **Table 6**). In pairs that are missing a number in the table, the dependencies are not known because none of the participants of the test groups had the pair in their individual BNs. **Table 6**, for instance, shows that although mariners believe that there is dependency between lack of VTS and lack of situational awareness, none of the researchers believe in that connection. What **Table 6** does not show is the mariners and researchers' beliefs regarding the causality links between the nodes (i.e. direction of the edges). Interested readers are referred to the complete report of the study [15], in where the direction of the edges by each group is discussed more comprehensively.

Table 6: The dependencies between the most frequent nodes in all accident reports. The number of connections has been divided with the number of BNs that have both nodes in them. The empty spaces in the table indicate that there are no BNs that include both nodes

	Bad visibility	Inadequate communications	Faulty practices	Hazardous natural environment	Mistake	Inadequate route plan	Inappropriate use of navigational equipment	Wrong steering decision	Lack of situational awareness	Fairway	Lack of VTS	Lack of guidelines	Grounding
Bad visibility		0,0	-	-	-	0,3	0,0	-	0,5	-	0,0	0,0	0,3
Inadequate communications	0,6		1,0	0,0	1,0	0,3	0,2	0,6	1,0	0,0	0,7	0,3	0,2
Faulty practices	0,0	0,7		0,5	0,5	0,4	0,6	0,2	0,3	0,7	0,5	1,0	0,2
Hazardous natural environment	0,0	0,2	0,5		1,0	0,7	0,0	0,5	0,0	-	-	0,0	0,4
Mistake	0,7	0,4	0,2	0,5		0,0	-	1,0	1,0	0,3	0,0	0,0	0,8
Inadequate route plan	0,3	0,0	1,0	0,0	1,0		0,4	0,5	0,6	0,0	0,2	0,7	0,2
Inappropriate use of navigational equipment	0,3	0,4	0,6	0,3	0,3	0,5		0,7	1,0	0,0	0,0	0,7	0,2
Wrong steering decision	-	0,7	0,5	0,0	1,0	-	-		0,8	0,0	0,0	0,2	0,7
Lack of situational awareness	-	0,3	0,5	0,0	1,0	-	0,3	-		0,8	1,0	0,0	0,7
Fairway	0,0	0,2	0,0	-	0,4	-	0,0	-	1,0		0,3	0,0	0,0
Lack of VTS	0,0	0,8	0,3	0,0	0,2	0,0	-	-	0,0	0,0		0,0	0,1
Lack of guidelines	0,0	0,7	0,2	0,4	0,0	0,7	0,4	0,5	0,0	0,0	0,0		0,0
Grounding	0,6	0,5	0,1	0,8	0,9	0,8	0,4	0,8	0,5	0,3	0,2	0,3	

Researchers (below the diagonal) / **Mariners** (above the diagonal)

4. DISCUSSION

The results of the study show that although in general neither of the groups have superiority over the other with regard to the extracted knowledge, there are some categories that are more frequently chosen by specific group. Mariners chose more often the causes related to navigation and the actions of the crew, while the researchers tend to see more organizational and environmental related causes. This was somehow expected as for instance the researchers may not know the importance of missing sea maps or inadequate route plan whereas mariners have personal experience on the matter. Therefore, causes like *mistake*, *wind*, and *fatigue* that are not related to navigation and thus are easier to find in the reports by people with no nautical background are more frequent in researchers' answers. Additionally, the researchers often connected darkness to bad visibility, while none of the mariners had darkness in their answers. The mariners may think that darkness should not affect the safety because it is unavoidable and vessels normally have enough equipment like radar to sail safely in darkness. There are also some categories that were chosen by the majority of the mariners but by none of the researchers in certain reports. For instance in the grounding of MV EMSRUNNER *inadequate route plan* was thought to have been a cause of accident by all mariners and in the grounding of MS PAULINE RUSS by half of the mariners but by none of the researchers. Also in the grounding of MS SUPERFAST VII the majority of mariners thought that *wrong steering decision* was one of the causes that led to the grounding. It was said in the answers that the use of tugboats was inadequate or the tugboats were not used correctly. None of the researchers thought that wrong steering decision caused the accident. These differences may be caused by the lack of navigational knowledge among the researchers as, for instance, the effect of tugboats may not have been directly pointed out in the report by the investigators thus making it harder for the researchers to see it as a contributing factor. This suggests that there are some causes that are mentioned in the storyline of the event but have not been investigated further in the accident reports; thus non-mariners are less likely to recognize them as contributing factors in the accident.

From another perspective, by looking at the results from an accident theory point of view like Reason's Swiss-Cheese [16], it is observable that the mariners have more "active failures" in their BNs whereas the categories that the researchers have more frequently in their BNs are "latent conditions". This means that the mariners are more concentrated in the person approach of the human factors than the researchers. On the other hand the mariners did also have latent conditions in their BNs but the way that they structured their BNs shows that they know latent conditions do not cause the accidents by themselves and often the crew is also responsible for the groundings. Of course the researchers also had categories from active failures in their BNs but not as much as the mariners. The mariners may know better from personal experience the types of active failures that cause accidents and how accidents can be avoided with different actions. The researchers may not be familiar with the rules and practices and if they are not mentioned in the accident reports clearly as accident causes the researchers may not know if all good-seamanship practices were followed.

On the other hand, the comparison of the whole BNs by calculating distances between the BNs shows that the consistency within the groups are not on average always more than between the groups. There were some instances where the distance in one group was higher than the distance between the groups. There are also many instances where there is a BN in one group that has smaller distances with the BNs of the other groups than with its own groups. Since the distances do not show us the differences in individual nodes and edges and merely show the differences in the networks as a whole, results of the distance measurement suggest that the two groups are not necessarily as different as what the comparison of the nodes showed when the whole networks are compared instead of individual nodes.

Despite of the above discussion and drawn results, there are uncertainties involved in the study that need to be mentioned and taken into account. Due to the lack of marine experts that were willing to participate in the study, the sample size in the study is limited. The sample size of each group should be at least ten to get test statistics that are good enough for statistical testing [17]. With only six observations in total, the statistical reliability of the study is not high enough. This is unfortunate though unavoidable because the study requires plenty of the participants' time. With such small sample size more accurate assumptions about the differences in the nodes should not be made. With bigger sample sizes the results could be analyzed statistically to get a better view on the differences between the groups. Now the variance of the results is too big to get reliable results on whether there is a significant difference between the two groups. Only the causes that were clearly picked by only one group can be said to have differential effects between the groups. Moreover, the inconsistency in the years of experience between the test groups may have introduced unknown uncertainties to the results. Therefore, the years of experience is certainly a factor that has potential for further studies in this regard.

In addition, different interpretation of the categories (nodes), like *inappropriate communication* that might be outside communication or bridge communication, might result in different ways of connecting the nodes with edges. Therefore, it is not easy to accurately say if the frequency of the edges varied between the groups' BNs. One reason for the differences may be that the causalities can be interpreted differently in building BNs. The most significant differences in the edges are the differently oriented edges. For instance in the accident report of the grounding of MS SUPERFAST VII, although there were no connection mentioned in the report between two causes as *wind* and *lack of guidelines*, there were two differently directed edges between the two nodes in the collected answers. It was said in the accident report that "*The instructions did not contain harbor maneuvering in a storm, which resulted in a defective estimation of the wind effect.*" and "*The wind limit had been exceeded*" [18]. The first sentence may imply that there should be an edge from lack of guidelines to hazardous natural environment as wind, which may come from an interpretation as: "Because of the lack of guidelines about wind limits the wind becomes a problem, whereas if there had been adequate guidelines the wind would not have been a problem." Two of the researchers and none of the mariners chose this combination to their BNs. Two other researchers though chose the opposite direction for this edge. These participants may have interpreted the interdependency differently as: "The high winds in a harbor area create a need for guidelines and the lack of guidelines creates a dangerous situation." Therefore, future studies needs to cover causality interpretation aspect more closely, either by providing a unique interpretation for each possible link, or by in-deep interviews regarding the implemented interpretation for each link.

5. CONCLUSION

In general, although there is no superiority between the groups for reviewing the accident reports with regard to the extracted knowledge, it seems that the researchers cannot identify the same BNs as the mariners and vice versa. There are clear differences between the two test groups of the study. Mariners chose more often the causes related to navigation and the actions of the crew, while the researchers tend to see more organizational and environmental related causes. Therefore, as was expected, the causal network of the same accidents varied both within and between the groups. Even with a finite set of possible categories none of the participants identified identical BNs from the accident reports of the study. From the study we can conclude that there are differences in the mathematical models that different individuals build from existing accident reports.

Acknowledgements

This study was conducted as a part of "Minimizing risks of maritime oil transport by holistic safety strategies" (MIMIC) project. The MIMIC project is funded by the European Union and the financing comes from the European Regional Development Fund, The Central Baltic INTERREG IV A Programme 2007-2013; the City of Kotka; Kotka-Hamina Regional Development Company (Cursor Oy); Centre for Economic Development, and Transport and the Environment of Southwest Finland (VARELY).

Merenkulun säätiö is also appreciated for providing the financial support for attending the PSAM12 conference to present this paper.

References

[1] Mazaheri A., and Ekwall D. (2009) "*Impacts of the ISPS code on port activities: a case study on Swedish ports*", World Review of Intermodal Transportation Research**,** Vol.2, No.4, pp.326-342, doi: 10.1504/WRITR.2009.026211

[2] Aven T. (2013) "*A conceptual framework for linking risk and the elements of the data-information-knowledge-wisdom (DIKW) hierarchy*", Reliability Engineering System Safety Vol.111, pp.30–36

[3] Mazaheri A., Montewka J., Kujala P. (2013) "*Modeling the risk of ship grounding - A literature review from a risk management perspective"*, Published online in WMU-Journal of Maritime Affairs, doi: 10.1007/s13437-013-0056-3

[4] Kristiansen S. (2010) "*A BBN approach for analysis of maritime accident scenarios*" In: Ale BJM, Papazpglou IA, Zio E (eds) ESREL. Taylor & Francis, Rhodes

[5] Tirunagari S., Hänninen M., Ståhlberg K., Kujala P. (2012) "*Mining Causal Relations and Concepts in Maritime Accidents Investigation Reports*", International Conference cum Exhibition on Technology of the Sea. Visakhapatnam: IMU

[6] Grech M. R., Horberry T., Smith A. (2002) "*Human error in maritime operations: analyses of accident reports using the Leximancer tool*", Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting. 46, pp. 1718-1722. Baltimore: Human Factors and Ergonomics Society. doi:10.1177/154193120204601906

[7] Mazaheri A., Sormunen O. V. E., Hyttinen N., Montewka J., Kujala P. (2013) "*Comparison of the learning algorithms for evidence-based BBN modeling – A case study on ship grounding accidents*", Proceedings of the Annual European Safety and Reliability Conference (ESREL), pp. 193-200, September 30th - October 2nd, Amsterdam, the Netherlands; ISBN 978-1-138-00123-7

[8] Johnson C. W., Botting R. M. (1999) "*Using Reason's Model of Organisational Accidents in Formalising Accident Reports*" Cognition, Technology & Work, Vol.1, pp.107-118

[9] Lekberg A. (1997) "*Different approaches to incident investigation-how the analyst makes a difference*", System Safety Society Conference, Hazard Prevention 33:4

[10] Dalkey N. C. (1969) "*The Delphi Method: An experimental study of group opinion*" Santa Monica, California: The Rand Corporation.

[11] Rothbaum A. R. (2000) "*Human Error and Marine Safety*", National Safety Council Congress and Expo. Orlando

[12] McCafferty D., Baker C. (2006) "*Trending the causes of marine incidents*" ABS Technical Papers 2006, 1-9.

[13] de Jongh M., Druzdzel M. J. (2009) "*A Comparison of Structural Distance Measures for Causal Bayesian Network Models*", In: M. Klopotek, A. Przepiorkowski, S. Wierzchon, & K. Trojanowski (Eds.), Recent Advances in Intelligent Information Systems, Challenging Problems of Science, Computer Science series (Vol. 1, pp. 443-456). Warsaw, Poland: Academic Publishing House EXIT.

[14] Acid S., de Campos L. M., Fernández-Luna J. M., Rodríguez S., Rodríguez J. M., Salcedo J. L. (2004) "*A comparison of learning algorithms for Bayesian networks: a case study based on data from an emergency medical service*", Artificial Inteligence in Medicine, 30, 215-232.

[15] Hyttinen N. (2013) "*The effect of experience on knowledge extraction from accident reports*", Master's thesis, School of Science, Aalto University, Espoo, Finland, p.80

[16] Reason J. (1990) "*Human Error*", Cambridge University Press, Cambridge, UK.

[17] Mellin I. (2006) "*Testejä laatuasteikollisille muuttujille*", Retrieved 07 09, 2013, from math.aalto.fi/opetus/sovtoda/oppikirja/Testit.pdf

[18] Safety Investigation Authority (2004) "*MS Superfast VII, grounding off Hanko on 12.11.2004*", Report No. B7/2004M