# The Effects of Causation Probability on the Ship Collision Statistics in the Gulf of Finland

## M. Hanninen & P. Kujala
*Helsinki University of Technology, Helsinki, Finland*

ABSTRACT: In this study the marine traffic safety in the Gulf of Finland is studied by examining the collision probability estimates in a heavily used crossing area. In a commonly applied approach for estimating the probability of collision accidents, so-called number of collision candidates is multiplied with a so-called causation probability. In this study a Bayesian network model for the causation probability estimation is applied with different parameter values in order to examine the effects of weather and human factors on collision probability in the crossing of Helsinki-Tallinn traffic and vessels navigating east- or westbound. The results show that the probability of collisions is very sensitive to the causation probability value and it should be modelled with great care to obtain reliable results.

## 1 INTRODUCTION

The Gulf of Finland is a sensitive geographical area. The Baltic Sea, including the Gulf of Finland, has been categorized as a Particularly Sensitive Sea Area (PSSA) by the International Maritime Organization (IMO, 2005). Maritime traffic is continuously increasing in the Gulf of Finland. Especially the increasing number of oil tankers is raising concern in the coastal countries. Russia is building new oil terminals, and the annual oil transports via the Gulf of Finland are estimated to increase even up to 250 millions of tons by 2015 (Finnish Environment Institute 2007). The increasing maritime traffic increases the risks of accidents, which could lead to oil spills. An oil disaster would most probably have serious effects on the Gulf of Finland ecosystem (Ihaksi et al. 2007).

Based on maritime accident statistics, groundings and collisions are the dominant accident types in the Gulf of Finland (Kujala et al. 2009). A commonly applied approach for estimating the probability of collisions or groundings in maritime traffic was defined by Fujii et al. (1971, 1974) and Macduff (1974). In this approach, the number of ships that would collide or run aground, if no evasive manoeuvres are made is calculated first. In the calculations it is assumed that the ships are sailing "blindly" in the waterway. This so-called number of collision candidates depends on the properties of ship traffic such as geometric traffic distribution on the studied waterway and ship sizes and speeds. In order to estimate the potential number of collisions or groundings, the number of collision candidates is then multiplied by the probability of not making evasive manoeuvres, so-called causation probability, which is conditional on the blind navigation assumption. The causation probability thus quantifies the proportion of cases when an accident candidate ends up grounding or colliding with another vessel. This approach for estimating the potential number of collisions or groundings can be expressed as

$$P = N_a \times P_C \tag{1}$$

where $N_a$ = the number of collision or grounding candidates; and $P_C$ = causation probability, i.e. the probability of not making evasive manoeuvres.

Not making an evasive manoeuvre while being on a collision or grounding course can be a result of a technical failure such as failure of steering system or propulsion machinery, human failure, or environmental factors. Technical failure was reported as the primary reason of the accident in 9.4 % of collision and grounding accidents in the Gulf of Finland, and in 25 % of the cases the primary reason had been conditions outside the vessel (Kujala et al. 2009). Human failure has been commonly stated as the most typical cause group of marine traffic accidents: different studies have shown that 43 % - 96 % of the accidents had been caused by humans (Grabowski et al. 2000, Hetherington 2006, Rothblum 2006, Kujala et al. 2009).

Causation probability values for crossing encounters in the literature have varied between $6.83 \cdot 10^{-5}$

– 6.00 · $10^{-4}$ (Macduff 1974, Fujii 1983, Fowler & Sørgård 2000, Otto et al. 2002, Rosqvist et al. 2002). The values have been either general values on some sea area, or reflecting certain ship types or conditions. In good visibility within VTS zone, Fowler & Sørgård (2000) estimated a causation probability of 6.83 · $10^{-5}$, and in poor visibility the value was 4.64 · $10^{-4}$. For collisions in the Gulf of Finland within VTS zone where at least one of the colliding vessels was a tanker, Rosqvist et al. (2002) estimated the value to be 5.1 − 6.0 · $10^{-4}$, depending on the other ship type.

In the earliest collision probability estimations the causation probability was estimated based on difference between the registered number of accidents and the estimated number of collision candidates (Fujii 1971, 1974; Macduff 1974). Applying a causation probability value derived from a study in another sea area or estimating it based on the difference in accident statistics and geometrical probability may save some effort, but then the actual elements in accident causation are not addressed at all, as opposed to constructing a model for the estimation. Getting a numerical value for the probability of not making an evasive manoeuvre is only one outcome of a model, the acquired model structure itself and the dependencies of the parameters may be at least equally important.

Risk analysis tools such as fault tree analysis have been used in modelling the causation probability (e.g., Pedersen 1995). In 2006, utilization of Bayesian networks in Step 3 of the Formal Safety Assessment was suggested in a document submitted by the Japanese agency for maritime safety to the IMO Maritime Safety Committee (2006). Bayesian networks are directed acyclic graphs that consist of nodes representing variables and arcs representing the dependencies between variables (e.g. Jensen 2007). Each variable has a finite set of mutually exclusive states. For each variable A with parent nodes $B_1,…, B_n$ there exist a conditional probability table $P(A \mid B_1, …, B_n)$. If variable A has no parents it is linked to unconditional probability $P(A)$. For identifying the relevant nodes and the dependencies between nodes, and constructing the node probability tables, both hard data and expert opinions can be used and mixed. Bayesian networks can also be used as an aid in decision-making under uncertainty. Bayesian networks have been applied in causation probability estimation such as in the maritime traffic risk assessment software GRACAT (Friis-Hansen & Simonsen, 2002) and in Øresund sound risk assessment (Rambøll, 2006).

The study described in this paper is a part of a cross-disciplinary approach for minimising the risks of maritime transport in the Gulf of Finland, where, based on growth predictions, the maritime traffic in the Gulf of Finland in the year 2015 is modelled and the accident risk, the direct environmental effects and the risk of environmental accidents are evaluated, and the effects of national and international legislation and other management actions are modelled (Klemola et al. 2008). In the previous work ship collision probabilities for two locations in the Gulf of Finland were estimated by applying causation probability values derived from literature (Kujala et al. 2009). This paper describes the application of a Bayesian network model for the causation probability modelling as a part of collision probability estimation for the traffic in a crossing area in the Gulf of Finland. The network is utilized for studying the effects of weather, human factors, and extra vigilance on the collision probability.
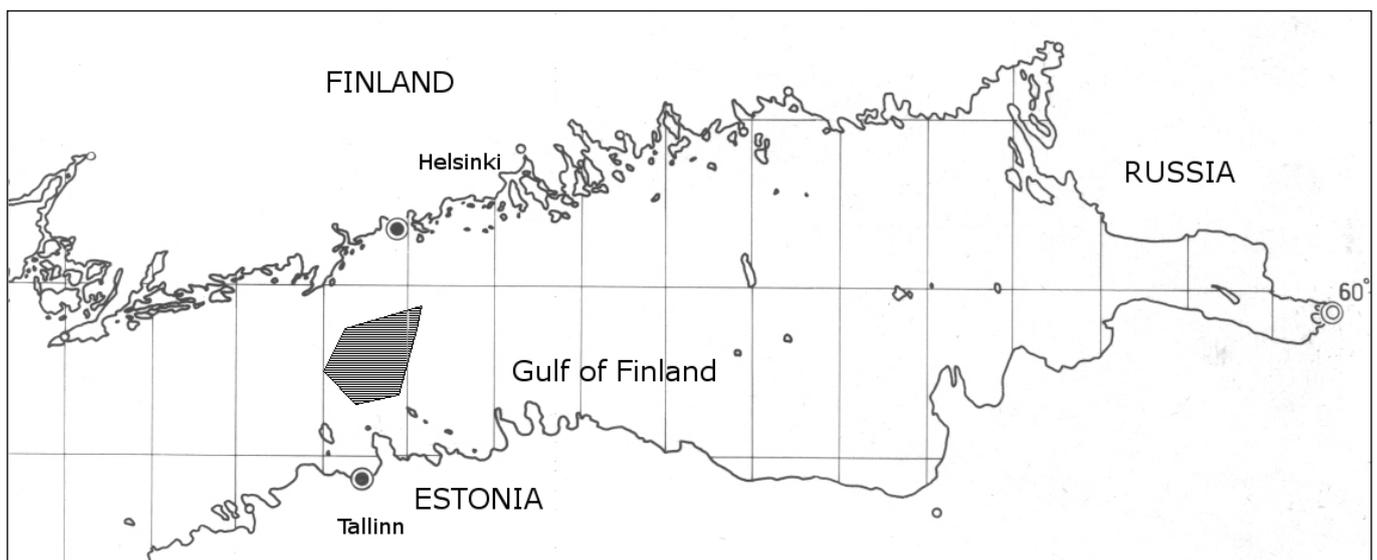


Figure 1. The studied crossing area between Helsinki and Tallinn marked in grey on the map.

## 2 STUDIED AREA: CROSSING BETWEEN HELSINKI AND TALLINN

The studied location (Fig. 1) is one of the highly trafficked crossings of waterways in the Gulf of Finland. In this area the vessel traffic between Helsinki and Tallinn is crossing the main route of the Gulf of Finland, i.e. vessels heading to and from Russia and eastern ports of Finland. Based on AIS records, in July 2006 there had been 2122 ships navigating north- or southbound, majority being fast ferries or passenger ships, and 2303 ships heading to and from eastern part of Gulf of Finland in July 2006 (Kujala et al. 2009). According to accident statistics, one collision of ships had been reported to occur in the area during six year period (Kujala et al. 2009).

## 3 MODEL USED FOR GEOMETRIC PROBABILITY

The number of collision candidates in the studied area during one summer month was estimated with a model presented by Pedersen (1995), which followed the concept introduced by Fujii (1971). The number of collision candidates in a time period was calculated as

$$N_a = \sum_i \sum_j \iint_{\Omega(z_i, z_j)} \frac{Q_{1i} \cdot Q_{2j}}{V_i^{(1)} \cdot V_j^{(2)}} \cdot f_i^{(1)}(z_i) f_j^{(2)}(z_j) \cdot V_{ij} D_{ij} dA \Delta t \quad (2)$$

where $N_a$ = the number of collisions if no evasive manoeuvres were made; $i$ and $j$ = ship classes of the colliding vessels; $Q_{1i}$ = the number of class $i$ vessels at waterway 1 in time unit; $Q_{2j}$ = the number of class $j$ vessels at waterway 2 in time unit; $V_i^{(1)}$ = the average velocity of class $i$ vessels at waterway 1; $V_j^{(2)}$ = the average velocity of class $j$ vessels at waterway 2; $f_i^{(1)}$ = the lateral distribution of traffic in waterway 1; $f_j^{(2)}$ = the lateral distribution of traffic in waterway 2; $V_{ij}$ = the relative velocity of ships depending on velocities and meeting angle; $D_{ij}$ = so-called geometrical collision diameter depending on vessel lengths, beams and velocities; and $\Delta t$ = time period under review.

The parameters of the collision candidate model were based on analysis of AIS records from the studied area in July 2006. The lateral distributions were approximated with normal distributions whose parameters were based on AIS records. For the calculations the vessels were grouped into five ship classes: passenger ships, cargo vessels, tankers, high speed crafts (HSCs), and other ships. Each class was divided into four size groups: length less than 100 metres, length at least 100 but less than 200 metres, length at least 200 m, and length unknown for which the average values of length and width of the ship class in question were used. The angle between crossing ships had been varying at the crossing point, so the average angle of arrival of each ship class from each approach direction was used in the calculations.
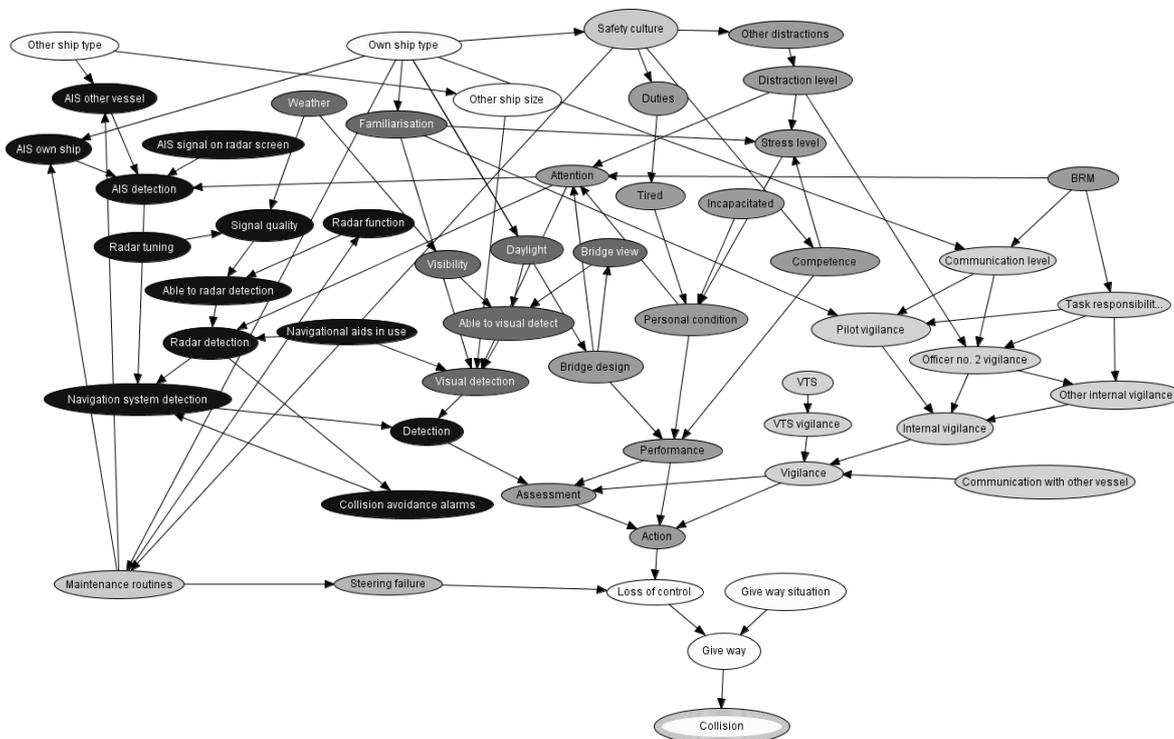


Figure 2. The applied Bayesian network structure for causation probability adapted from (Det Norske Veritas 2003, 2006).

Table 1. Causation probability network node states whose probability were set to 1.0 in good and poor environmental and human performance conditions

| Node | Environmental conditions | | Human performance | |
|---|---|---|---|---|
| | good | poor | good | poor |
| Daylight | day | night | - | - |
| Visibility | > 1 nm | < 1 nm | - | - |
| Weather | good | storm | - | - |
| Attention | - | - | high | low |
| Communication level | - | - | beyond standard | sub-standard |
| Communication with other vessel | - | - | yes | no |
| Competence | - | - | high | low |
| Distraction level | - | - | low | moderate |
| Duties | - | - | normal | extreme |
| Stress level | - | - | low | high |
| Tired | - | - | no | yes |

## 4 MODEL USED FOR CAUSATION PROBABILITY

The applied Bayesian network model for estimating the causation probability, i.e. the probability of not making evasive manoeuvres, was based on fragments of a collision model network in the Formal Safety Assessment of large passenger ships (Det Norske Veritas 2003) and a grounding model in the FSA of ECDIS chart system (Det Norske Veritas 2006). The network estimated the probability of a collision given that two ships were on a collision course, one ship had lost control and the other ship did not give way. The network included parameters related to navigational aids, conditions, safety culture, personnel factors, management factors, other vigilance, and technical reliability. The network reflected the following events for making an evasive manoeuvre while on collision course. At first the Officer On Watch (OOW) had to detect the dangerous situation either visually or with navigational aids. Detection was influenced by parameters related to external and internal conditions as well as attention. After the detection, OOW had to make a correct assessment of the situation, which was influenced by OOW's performance level. Situation might have also been assessed correctly even without OOW's detection if other vigilance such as a pilot or VTS operator was present to detect the danger. If situation was assessed correctly, OOW had to make an avoiding act. If control was lost because of either wrong or no action or steering failure, the collision might have still been avoided if the other ship gave way. The network was modified so that it was suitable to be applied to an analysis including multiple ship types. The network structure can be seen in Figure 2.

Most of the probability values related to the Bayesian network parameters were derived from the original models and had been mostly based on ex-

pert judgment. Ship type distributions in the waterways of the studied area were obtained from AIS-data. The probability distributions of "Weather" states were based on Finnish Meteorological Institute's statistics on the average number fog days at Isosaari in July during 1961-2000, the average number of storm days at Finnish sea areas in July during 1990-2008 thinned by the average portion of storm observations from the Gulf of Finland in 2006-2007, and the average number of strong wind days at Isosaari in July during 1961-2000 (Finnish Meteorological Institute, 2008). The daylight distribution describing the probability of a ship navigating in the dark, conditional on ship class, was based on AIS information and sunrise and sunset times at the studied location at 15.7.2006. The probability of "VTS" state "yes" was set to 1.0 because the studied area is monitored by VTS stations.

The effects of conditions outside the vessel and factors related to human performance on collision probability were studied by constructing scenarios describing different environmental conditions and/or factors related to human performance. The states of the nodes, the probability of which was set to 1.0 in the different environmental and human performance conditions are shown in table 1. For example, the environmental conditions were defined as "poor", if all of the following probabilities in the network were equal to 1.0:

– P(Weather = "storm")
– P(Visibility = "< 1 nm")
– P(Daylight = "night")

Causation probability was estimated for scenarios where 1) there was no evidence on any of the network parameters; 2) it was known that environmental conditions were "good" and the factors related to human performance were "good"; 2) it was known that environmental conditions were "good" and the factors related to human performance were "poor"; 3) it was known that environmental conditions were "good" but there was no information on other parameters; 4) it was known that environmental conditions were "poor" and the factors related to human performance were "good"; 5) it was known that environmental conditions were "poor" and the factors related to human performance were "poor"; 6) it was known that that environmental conditions were "poor" but there was no information on other parameters; 7) it was known that the factors related to human performance were "good" but there was no information on other parameters; 8) it was known that the factors related to human performance were "poor" but there was no information on other parameters. In addition, causation probability was estimated for situations where 10) there was no extra vigilance present for detecting the danger; and 11) danger was detected by VTS or other internal vigi-

lance. In situations 10 and 11 there was no evidence on any other parameters than the node "Vigilance". The network was built and the probability calculations were performed with Bayesian network software Hugin.

## 5 RESULTS OF THE ANALYSIS

General causation probability for the studied location and traffic, meaning that there was no additional knowledge on the network parameters other than the default conditional probabilities of the network, was estimated to be $2.70 \cdot 10^{-4}$. When multiplied by the number of collision candidates, the resulting number of collisions in one month was $1.64 \cdot 10^{-2}$ which equals a return period of 61 months. If it was certain that the danger had been detected by extra vigilance, causation probability estimate was $2.58 \cdot 10^{-4}$ producing return period of 64 months. On the other hand, if there was no extra vigilance, causation probability was $3.74 \cdot 10^{-4}$ and the collision return period decreased to 44 months.

Tables 2 and 3 present causation probability and the expected number of collisions in a month estimates with different environmental and human factor conditions. The lowest collision probability in these scenarios was acquired in good environmental conditions with good human factors, and the collision probability was highest when both the environmental and human factor conditions were poor.

Table 2. Results of causation probability estimation with different scenarios related to environmental and human factor conditions

| Environmental conditions | Human performance | | |
| --- | --- | --- | --- |
| | Good | Poor | No evidence |
| Good | 2.56E-04 | 4.27E-04 | 2.68E-04 |
| Poor | 2.94E-04 | 1.97E-03 | 7.01E-04 |
| No evidence | 2.56E-04 | 4.44E-04 | 2.70E-04 |

Table 3. Estimates of the number of collisions in a month with different scenarios related to environmental and human factor conditions

| Environmental conditions | Human performance | | |
| --- | --- | --- | --- |
| | Good | Poor | No evidence |
| Good | 1.55E-02 | 2.59E-02 | 1.63E-02 |
| Poor | 1.78E-02 | 1.19E-01 | 4.25E-02 |
| No evidence | 1.55E-02 | 2.69E-02 | 1.64E-02 |

## 6 CONCLUSIONS

The effects of weather and factors related to human performance on the collision probability were studied using a Bayesian network model for estimating the probability on not making an evasive manoeuvre while ships were on a collision course in crossing area between Helsinki and Tallinn in the Gulf of Finland. The general causation probability was estimated to be $2.70 \cdot 10^{-4}$, which is about the same order than the values found in literature. With this causation probability, the return period of collisions in the crossing area between Helsinki and Tallinn was estimated to be 5 years. According to statistics, one collision had occurred in the area in 6 years so it could be stated that the model reflected well the actual situation. However, it should be noted that it is hard to compare the results to statistics since analyzed time interval should be long but the traffic would have to remain constant. The return periods were estimated based on one summer month traffic data. The traffic in the area is very different during in winter period. Thus the effects of winter should also be included in modelling in the future.

According to the applied model, if human performance factors were poor and the ship would be sailing in difficult conditions at dark, the probability of a collision in the studied area would be almost eight times as big as in good environmental and human performance conditions. If just the difference in human performance is examined, the collision probability with poor human performance factors would be almost twice the probability in good human performance conditions. This evaluation shows that the validity of the network parameters is important in order to produce realistic estimates of collision probabilities. In the future expert judgment and ship simulator studies will be utilized in order to validate the model to the traffic and conditions in the Gulf of Finland. With a valid model the effects of possible risk control options on collision probabilities can be evaluated and the model can be used as an aid in decision-making.

All theoretical analysis completed in this document is based on data of only one month, July 2006. The amount of traffic is largely dependent on the season as well. Naturally this also means that the comparison with the accident statistics and theoretical model using only data from one month raises some concerns. This paper can, however, be considered as a good start to more profound analysis of the causation probability in the area, which should be conducted on monthly basis covering the whole year and based on data from other months as well.

## REFERENCES

Det Norske Veritas. 2003. *Formal Safety Assessment – Large Passenger Ships, ANNEX II*.
Det Norske Veritas. 2006. *Formal Safety Assessment of Electronic Chart Display and Information System (ECDIS)*. Technical Report No. 2005-1565, rev. 01.
Finnish Environment Institute. 2007. *Oil transportation in the Gulf of Finland through main oil ports – Oil transportation*

*in years 1995-2005 and estimated development by year 2015.*

Finnish Meteorological Institute. (n.d.) Weather statistics. FMI www pages (online) [cited 8.12.2008] Available in www-format: <URL:http://www.fmi.fi/>.

Fowler, T. G. & Sørgård, E. 2000. Modeling Ship Transportation Risk. *Risk Analysis* 20(2): 225-244.

Friis-Hansen, P. & Simonsen, B.C. 2002. GRACAT: software for grounding and collision risk analysis. *Marine Structures* 15(4): 383-401.

Fujii, Y. & Shiobara, R. 1971. The Analysis of Traffic Accidents. *Journal of Navigation* 24(4): 534-543.

Fujii, Y., Yamanouchi, H. & Mizuki, N. 1974. Some Factors Affecting the Frequency of Accidents in Marine Traffic. II - The Probability of Stranding and III - The Effect of Darkness on the Probability of Collision and Stranding. *Journal of Navigation* 27(2): 239-247.

Fujii, Y. 1983. Integrated Study on Marine Traffic Accidents. *IABSE Colloquium on Ship Collision with Bridges and Offshore Structures, Copenhagen* 42: 91-98.

Grabowski, M., Merrick, J. R. W., Harrald, J. R., Mazzuchi, T. A. & van Dorp, J. R. 2000. Risk Modeling in Distributed, Large-Scale Systems. *IEEE Transaction on Systems, Man, and Cybernetics – Part A: Systems and Humans.* 30(6): 651-660.

Hetherington, C., Flin, R. & Mearns, K. 2006. Safety in shipping: The human element. *Journal of Safety Research* 37(4): 401-411.

Ihaksi, T., Helle, I., Lecklin, T., Ryömä, R., Kokkonen, T. & Kuikka, S. 2007. Using biological knowledge and decisions of society in spatial prioritization of oil combating, *ICES CM 2007/C:02*.

IMO. 2005. *Revised Guidelines for the Identification and Designation of Particularly Sensitive Sea Areas (PSSAs),* Resolution A.982(24).

IMO. 2006. *Formal Safety Assessment. Consideration on Utilization of Bayesian Network at Step 3 of FSA,* Maritime Safety Committee 81st Session, MSC 81/18/1.

Jensen, F. V. & Nielsen, T. D. 2007. *Bayesian Networks and Decision Graphs, 2nd Edition.* New York: Springer.

Klemola, E., Kuronen, J., Kalli, J., Arola, T., Hänninen, M., Lehikoinen, A., Kuikka, S., Kujala, P. & Tapaninen. U. 2008. A cross-disciplinary approach to minimising the risks of maritime transport in the Gulf of Finland. *World Review of Intermodal Transportation Research,* in Press.

Kujala, P., Hänninen, M., Arola, T. & Ylitalo, J. 2009. Analysis of the marine traffic safety in the Gulf of Finland. Submitted manuscript.

Macduff, T. 1974. The probability of vessel collisions. *Ocean Industry* 9(9): 144–148.

Otto, S., Pedersen, P. T., Samuelides, M. & Sames, P. C. 2002. Elements of risk analysis for collision and grounding of a RoRo passenger ferry. *Marine Structures* 15(4): 461-474.

Pedersen, P. T. 1995. Collision and Grounding Mechanics. *Proceedings of WEMT'95*: 125-127. Copenhagen: The Danish Society of Naval Architects and Marine Engineers.

Rambøll. 2006. *Navigational safety in the Sound between Denmark and Sweden (Øresund); Risk and cost-benefit analysis.* Rambøll Danmark A/S.

Rosqvist, T., Nyman, T., Sonninen, S. & Tuominen, R. 2002. The implementation of the VTMIS system for the Gulf of Finland - a FSA study. *RINA International Conference Formal Safety Assessment*, London 18.-19.9.2002. The Royal Institution of Naval Architects (RINA): 151-164.

Rothblum, A. M. 2006. *Human error and marine safety.* Volume 4 in U.S. Coast Guard Risk-Based Decision-Making Guidelines. U.S. Coast Guard Research and Development Center, (online) [cited 8.12.2008]. Available in www-format: <URL:http://www.uscg.mil/hq/g-m/risk/e-guidelines/RBDM/html/vol1/ContV4.htm>