TRANSNAV
http://www.transnav.eu
the International Journal on Marine Navigation and Safety of Sea Transportation
Volume 7
Number 1
March 2013

DOI: 10.12716/1001.07.01.16

Correlation between the Ship Grounding Accident and the Ship Traffic – A Case Study Based on the Statistics of the Gulf of Finland

A. Mazaheri, J. Montewka & P. Kujala
Aalto University, Department of Applied Mechanics, Espoo, Finland

ABSTRACT: Ship traffic is one of the factors that is presented in almost all of the existing grounding models, and is considered as one of the affecting factors on the likelihood of grounding accident. This effect in grounding accident is mostly accepted by the experts as a common sense or simply by just generalizing the ship-ship collision cases to grounding accidents. There is no available research on the actual causal link between the ship traffic and grounding accident in the literature. In this paper, authors have utilized the statistical analysis on historical grounding accident data in the Gulf of Finland between the years 1989 and 2010 and the AIS data of the same area in year 2010, as the source of ship traffic data, to investigate the possible existence of any correlation between the ship traffic and the grounding accident. The results show that for the studied area (Gulf of Finland) there is no correlation between the traffic density and the grounding accident. However, the possibility of the existence of minor relation between the traffic distribution and grounding accident is shown by the result. This finding, however, needs further investigation for more clarification.

1 INTRODUCTION

From among the many factors that scholars have considered as affecting factors on the likelihood of grounding accident when they have modeled this type of accident, traffic is one of the factors that are presented in almost all of the existing models related to grounding accident; see e.g. [1-6]. Similar to the correlation of the ship traffic and frequency of ship-ship collision that is merely noted in the literature [7], it seems that this effect in grounding accident is mostly accepted by the experts as a common sense or simply by just generalizing the ship-ship collision cases to grounding accidents. There is no available study on the actual causal link between the traffic and grounding accident in the literature. It can be argued that one of the reasons behind the common belief of the existence of causal link between the traffic and grounding accident is that people assume when the traffic is more dense, the likelihood that the ships have to alter their courses to avoid collision and eventually ending up grounded is higher; however there is no statistical analysis on this common belief to either support or reject it. One of the problems that may rise when such doubtful beliefs become commonly accepted by the researchers is that the models that are developed to analyze a phenomenon will not be accurate enough and may not represent the reality; and if the model is used for risk management purposes it might result in ineffective or even wrong risk control options.

In this paper, it has been tried to test this common belief by finding the possible correlation between the two variables, using the statistical data of the actual grounding accidents happened in the Gulf of Finland (GOF) within 22 years (1989-2010).

The remainder of this paper is organized as follows: The data that is used for the analysis is presented in the next Chapter. The research methodology and the implemented algorithms for data analysis are explained and presented in Chapter 3; the results of the data analysis are presented in Chapter 4; followed by the discussion of the results in Chapter 5. The paper is concluded in Chapter 6.

2 DATA

In order to analyze the possible correlation between the traffic of ships and the grounding accidents, two different sources of historical data are used as: 1- HELCOM (Helsinki Commission) database regarding the ship accidents happened in the Baltic Sea area within the years of 1989 and 2010; 2- HELCOM AIS (Automatic Identification System) data on marine traffic in the GOF in year 2010.

2.1 *HELCOM Accident Database*

The first dataset that is utilized in this paper is HELCOM database regarding the ship accidents previously occurred in the Baltic Sea area, including Gulf of Finland, between the years of 1989 and 2010. The data consist of inputs such as the date and time of the accident, geographical coordinate of the accident, type of the accident, flag states of the involved ships, name of the involved ships, whether the accident caused any pollution, and type and amount of the possible pollution. From among all the available fields in the database, the only input values that have been utilized in this research are the type and the location of the accident. Although the database was not flawless, especially regarding the ship and cargo properties and crew competences, when it comes to the data useful for the purpose of this study, the only problem that was needed to be addressed seems to be the wrong recorded location of the accident that in some cases were reported in land areas.

The data were first filtered regarding the type of the accident, in which all the accidents that were recorded as grounding were chosen. Before utilizing the database, the database was filtered to limit the data to those accidents that have happened in the GOF, means limiting the coordination to 21.63^{o} E and 30.31^{o} E longitude, and to 58.90^{o} N and 60.89^{o} N latitude. There were in total 616 records of grounding accidents in HELCOM database for the years of 1989-2010, in which 123 of them were occurred in the Gulf of Finland based on the above geographical limitation. Next, the database was filtered to remove those accidents that have reported as occurred in land area. To do so, the accident points were visualized in GIS software over the map of the GOF and then those accident points that have located in the land area were spotted and manually removed from the database (Figure 1). From among 123 grounding accident records in the GOF, 11 were found as registered by wrong coordination. As the result, 112 grounding cases spotted as happened in the GOF between years 1989 and 2010, and have been used in the statistical data analysis for this paper. Here in this paper, the location of these 112 grounding accidents are referred as grounding points.

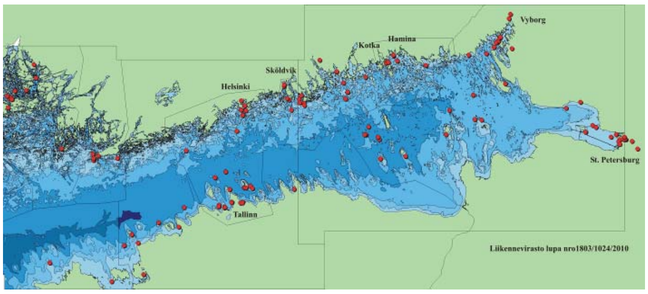

Figure 1. The location of the grounding accidents happened in the Gulf of Finland between 1989-2010

2.2 *HELCOM AIS Database*

The second source of historical data that is utilized in this study is the AIS data of the marine traffic in the GOF in 2010. The AIS data of the ship traffic has been used to extract the recent traffic properties of ships navigating in the GOF. The AIS data of the ships navigating in the Baltic Sea area were all reported and stored in the database of HELCOM. The database have input values such as IMO and MMSI numbers of the vessel, timestamp of the AIS record, geographical coordinate of the AIS record as latitude and longitude, main dimension of the vessel, speed over ground, course over ground, and rate of turn. However, the only input values that were utilized in this research were timestamps and geographical location of the ships. The AIS data received from HELCOM were filtered and sorted using the methodologies suggested in the literature [8,9].

3 METHODOLOGY

The methodology adopted in this paper is integrated into two main phases. The first phase is to define algorithms to detect the traffic definitions as Traffic Density and Traffic Distribution in the utilized AIS data; and the second phase was implementing statistical hypothesis testing in order to find whether there is any correlation between the traffic and the historical location of the grounding accident using the defined algorithms.

3.1 *Algorithm Definitions*

In the literature, the traffic of the ships was utilized in two ways of definitions as Traffic Density and Traffic Distribution (lateral distribution of the ships along the path). Traffic density is defined as the number of ships per unit area of the waterway within a desired time window [10, 11]. Since we are dealing with the AIS data of year 2010, the time window for this study is defined as a year. The algorithms to estimate the traffic density from the AIS data based on the above definition is shown in Figure 2. In general, the Gulf of Finland is divided into grid cells of size one by one nautical mile. Thereafter, the annual traffic density in the cells that has at least a grounding point inside is estimated. The annual traffic density is counted as the number of the ship tracks that passed through such cells over a year. By this way of algorithm definition, the speeds of the vessels do not affect the results as the linear interpolation between the available AIS data points will remove the effect from the interval of the AIS points that are varied depends on the instant speed of the vessel. Another variable is defined in the algorithm to keep the numbers of grounding points that are located inside the grid cell that the traffic density is estimated. This variable has later been used to find the correlation between traffic density and number of groundings.

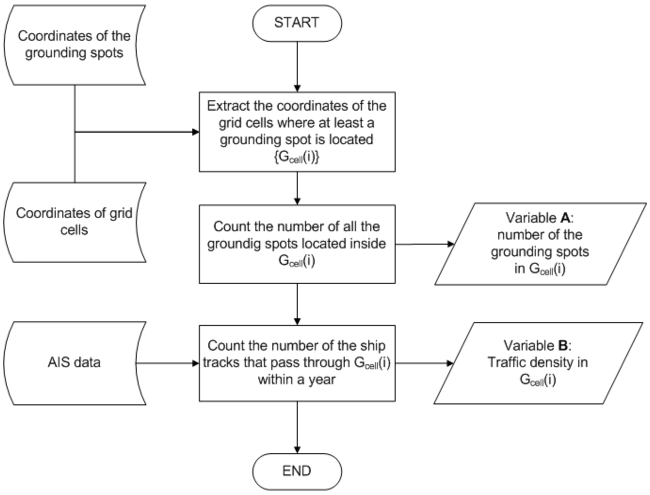

Figure 2. Algorithm to extract the traffic density from the AIS data

Lateral distribution of the ships or ship track distribution is the other way of defining the traffic of the ships when the probability of grounding is estimated in the literature; see e.g. [2-4, 12-16]. Nevertheless, there is no unique definition regarding where exactly along a path the distribution of the ships should be extracted. Obviously the extracted lateral distributions of the ships would be different depends on where they have been extracted; the closer to the shoal, the more corrected courses by the ships, thus the narrower or skewer distributions. In order to avoid confusion and also to simplify the algorithm for extracting the data, a definition of the ship distance from a grounding point is defined here. The utilized algorithm is shown in Figure 3. In general, the Gulf of Finland is divided into grid cells of size five by five nautical miles. The five nautical miles distance is estimated to be the distance that can be travelled by most of the merchant ships (excluding High Speed Light Crafts) in half an hour using the average speed of the ships navigating in the GOF. Thereafter, the distribution of the ship traffic in the cells that has at least a grounding point inside is estimated. The distribution is estimated in this way that when the track of a ship is passed through the defined grid cell, the distance of the ship on the border of the grid cell from the grounding point is calculated and stored in a variable. Another variable is also counting the number of grounding points inside the defined grid cell. Later on, the correlation between the distributions of the calculated distances and the number of the counted grounding points are analyzed based on the first and second moments of the obtained distributions. This means that it is assumed that a distance distribution of the ships from a grounding point can be acceptably described by mean and variance of the distribution.

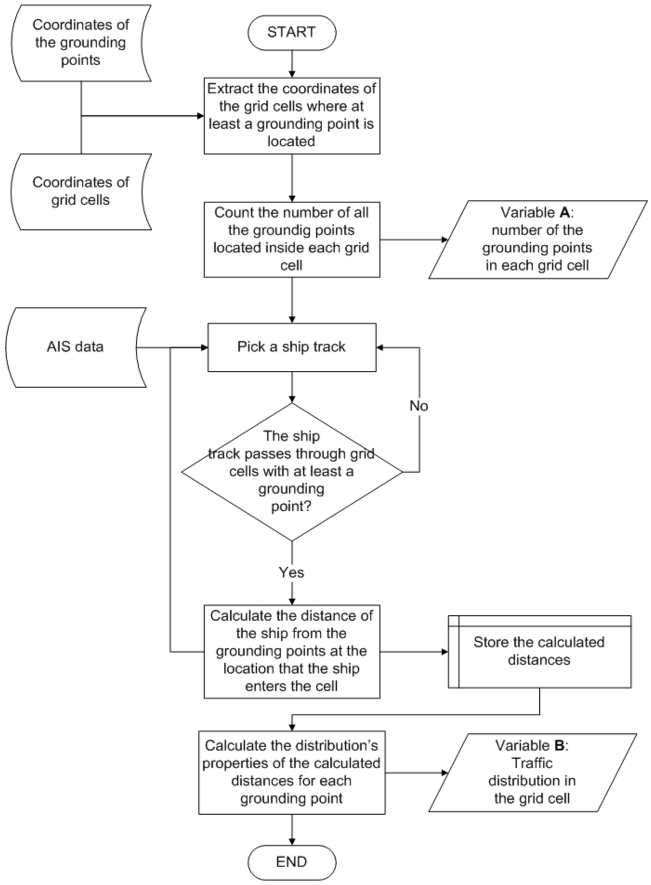

Figure 3. Algorithm to extract the traffic distribution from the AIS data

3.2 *Statistical Analysis*

One of the methods to find the statistical dependencies between two random variables is to find if any correlation exists between the two variables. Knowing about the correlation between two variables is specifically useful as it indicates a predictive relationship between the two variables, which can be exploited in practice. Nevertheless, the existence of statistical dependencies as correlation between two variables does not necessarily imply causal relation between the two variables. For more information regarding the correlation and causal relation the readers are refer to [17]. There are several coefficients that represent correlation dependencies between two random variables, which in this paper two coefficients as Pearson and Spearman have been utilized. Pearson coefficient (r) of two variables is defined as the covariance of the variables divided by the product of their standard deviations. Pearson coefficient gives a value between 1 and -1, which the exact value of 1 and -1 means there is a perfect linear correlation between the variables. The value of zero means that there is no linear correlation between the two variables. Nevertheless, "no linear correlation" cannot be interpreted as "absolutely no correlation" between the two variables; and still some sort of correlation in the form of nonlinear correlation, might be existing between the two variables. Although, there is still no unique method or coefficient that can reveal absolute correlation between the two variables, there are some coefficients like Spearman that can reveal some level of nonlinear correlation between two variables. Spearman coefficient (ρ) reveals if the relation of the two variables can be described using a monotonic function; thus it can grasp a degree of non-linear correlation between the two variables. Spearman coefficient gives value between 1 and -1, which the exact value of 1 and -1 means that there are no repeated data values, and each variable is a perfect monotonic function of the other.

Although Spearman and Pearson coefficients can somehow disclose the possible statistical dependencies between the two variables, one should be aware about their limitations and assumptions. For instance, Pearson coefficient is defined assuming the data are normally distributed, so in other cases it might be misleading; or Spearman coefficient is recommended when both variables are ordinal variables, or one is ordinal and the other is continues variable [17]. Therefore, since these assumptions are not perfectly matched with the limitation of our variables, some other methods like *Mutual Information* test should be used additionally. However, other tests like mutual information are not utilized in this paper and remained for the future research.

4 RESULTS

After calculating r and ρ coefficients for each sets of variables, the statistical significance of the results are tested assuming the null hypothesis (H_0) as "no-correlation" against alternative hypothesis (H_1) as "non zero correlation" by χ^2 test in significant levels of 95% (i.e. α = 0.05). The results are all shown in Table 1.

As is seen in Table 1, H_0 cannot be rejected for the traffic density and also for the mean value (mu) of the traffic distribution in 95% significant level. Therefore the validity of the null hypothesis is consistent with the resultant data. Thus, the existence of any correlation between the two variables as traffic density of the ships and the grounding accident is questionable. It should be noticed that the defined null hypothesis is a composite hypothesis; thus the trueness of the hypothesis cannot strictly verify that there is absolutely no correlation between the ship traffic and grounding accident; however it can strongly question its existence.

Table 1. Correlation between the number of groundings and traffic as density and distribution with two coefficients as Pearson (r) and Spearman (ρ)

Traffic Property	Coefficient	Correlation Value	P-value	Accepted Hypothesis
Traffic	r	0.0045	0.9654	H0
Density	ϱ	0.1102	0.2850	H0
Distribution	r	-0.0573	0.5714	H0
(mu)	ϱ	0.0093	0.9272	H0
Distribution	r	0.2462	0.0135	H1
(std)	ϱ	0.2418	0.0154	H1

One interesting point that can be seen from Table 1 is that the null hypothesis can be rejected for standard deviation (std) of the traffic distribution in 95% significant level. This might be the sign of slight correlation between the distributiveness of the ship traffic along the path and the grounding accidents. Although, the existence correlation is not very significant (less than 0.25), it has the potential for further investigation.

5 DISCUSSION

The result of this study shows that there is no significant correlation between the density of ship traffic and the grounding accidents, while it shows slight correlation between the distributiveness of the traffic and grounding accident.

Although this is an important and interesting result by its own as it is a counter claim for the currently existing common belief in the society, it should be used by caution and it needs further research for fully confirmation due to the available uncertainties. The three main sources of the uncertainties are the defined algorithms to extract the data regarding the traffic density and the ship traffic distribution, the issues regarding the utilized data, and the used statistical methods.

The algorithms defined in this research to extract the data based on the utilized definition of the ship traffic in the literature have a main issue, which is the size of the used grid cells. For the traffic density, the grid size of 1 by 1 nautical mile and for the traffic distribution 5 by 5 nautical miles are utilized. Therefore, the number of the grounding points that will be caught by these grid sizes may be changed if the grid size is increased or decreased. Besides, the number of the ships in the area may also be changed if the size of the grid is changed. Thus, the effect of the grid size on the result should be investigated by choosing different grid sizes. The question is the factors that affect the grid size. One important factor is the average speed of the traffic in the area, which in this study is believed to be neutralized by the utilized algorithm. The other factor might be the traffic congestion in the area, which may be neutralized by affecting the width of the waterway. However, one may argue that by affecting the width of the waterway the result is biased to the location. The counter argument would be that the nature of such research is in fact biased to the location of the previous grounding accidents; and in fact the grounding accident, in contrary to the collision accident, is very location dependent as it only may happen in the shallow water areas.

Nevertheless, the existence of any correlation between the grounding accident and the location (waterway) has a great potential for further investigation, especially since the slight correlation between the grounding accident and the distributiveness of the traffic can be seen as the effect of the width of the waterway on the traffic [18] and thus on the grounding accident.

Furthermore, the size of the ships navigating in the area may also affect the results, as in some definition of traffic density the size of the ships is an affecting factor [11]. Therefore, the algorithm should be further modified in order to take into account the size of the ship presented in the traffic.

One another source of uncertainty is that the AIS data utilized for this research represent the recent ship traffic in the area (year 2010), while the historical accident data were from the years 1989-2010. The current traffic does not necessarily represent the actual traffic density and distribution in the past. In fact, the ship traffic in the area has significantly increased during the past decades, due to the opening of the new ports in the area and also the economy growth of the neighboring nations [19]. Therefore, the result of this study needs to be verified using the AIS data of different years. Moreover, the AIS data used for this study covers the whole year of 2010; however, the studied area is normally covered with ice during the winter. The icy waterways may affect the traffic pattern in the area, which in this study is neglected. Thus, the effect of winter traffic on the result should also be investigated later.

The other matter is related to the hypothesis testing, where the null hypothesis can never be proven and it can only be "rejected" or "fail to reject". Failing to reject a null hypothesis does not mean that the null hypothesis is always true, rather is showing the null hypothesis is consistent with the resultant data; meaning that there is no enough evidence in the historical data to prove the opposite. Thus, although the existence of any possible correlation between the maritime traffic and the grounding accident is doubted by the result of this study, it certainly cannot be concluded as absolutely "no correlation". Besides, the utilized coefficients as Pearson and Spearman may not fully detect the existence of nonlinear correlation between the two variables. Therefore, implementing other methods, like mutual information test, seems useful in order to decrease uncertainty of the results.

The last but not the least matter is that whether "non-correlation can imply non-causation". The opposite statement as "correlation does not imply causation" is widely accepted between the statisticians [20]; however, "non-correlation implies non-causation" is still being discussed within the statisticians and other scholars [21]. Distinguishing a true causal relationship is very difficult and cannot be directly resulted from a correlation test. Therefore, even if the results of this research can be accepted as the proof for non-correlation between the density of ship traffic and grounding accident, still the causality link between these two should be investigated and discussed further.

6 CONCLUSION

It is shown in this research that there is no correlation between the ship traffic density and the grounding accidents, while there is slight correlation between the grounding accident and the traffic distributiveness.

The results are based on the historical grounding accidents that took place in the Gulf of Finland within the years 1989-2010 and the ship traffic of the same area in 2010. Thus, it is worth to highlight again that the obtained results are only valid for the studied area, and they cannot be generalized over other locations without further investigation.

There are some levels of uncertainty involved in the presented result, which are mostly related to the utilized algorithms to extract the required data from the databases. Some assumptions like the used grid cells should be tested against the different sizes in order to find the effect of the size of the grid cells in the final result. Besides, the effect of the winter traffic and size of the ships on the results, which are neglected in this study, have the potential of further investigation. More importantly, the non-existence of any causal link between the ship traffic and the grounding accidents cannot be merely concluded from the result of this research, and it needs further research and discussion.

AKNOWLEDGMENT

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).

REFERENCES

1. Fujii, Y. and R. Shiobara, The Analysis of Traffic Accidents. The Journal of Navigation, 1971. 24(4): p. 534-543.
2. Pedersen, P.T. Collision and Grounding Mechanics. in Proceedings of WEMT '95'. 1995. Copenhagen, Denmark: The Danish Society of Naval Architecture and Marine Engineering.
3. Simonsen, B.C., Mechanics of Ship Grounding, in Department of Naval Architecture and Offshore Engineering. 1997, Technical University of Denmark: Kongens Lyngby.
4. COWI, Risk Analysis for Sea Traffic in the Area around Bornholm. 2008, Danish Maritime Authority: Kongens, Lyngby.
5. Kristiansen, S., Traffic-Based Models, in Maritime Transportation-Safety Management and Risk Analysis. 2005.
6. Mazaheri A., et al., A Decision Support Tool for VTS Centers to Detect Grounding Candidates. TransNav - International Journal on Marine Navigation and Safety of Sea Transportation, Vol. 6, No. 3, pp. 337-343, 2012
7. Judson, B., Collision risk circumstances and traffic routeing in the approaches to the strait of Juan de Fuca, in 7th International VTS Symposium. 1992: Vancouver, Canada.
8. Goerlandt, F. and P. Kujala, Traffic simulation based ship collision probability modeling. Reliability Engineering and System Safety, 2011. 96(1): p. 91-107.
9. van Dorp, J.R. and J.R.W. Merrick, On a risk management analysis of oil spill risk using maritime transportation system simulation. Annals of Operations Research, 2009: p. 1-29.
10. Fujii, Y., et al., Some Factors Affecting the Frequency of Accidents in Marine Traffic: I- The Diameter of Evasion for Crossing Encounters, II- The probability of

Stranding, III- The Effect of Darkness of the Probability of Collision and Stranding. The Journal of Navigation, 1974. 27(2): p. 239-247.
11. Mazaheri, A. and J. Ylitalo, Comments on geometrical modeling of ship grounding, in 5th International Conference on Collision and Grounding of Ships (ICCGS). 2010: Espoo, Finland.
12. Otto, S., et al., Elements of Risk Analysis for Collision and Grounding of a RoRo Passenger Ferry. Marine Structures, 2002. 15: p. 461-474.
13. Karlsson, M., F.M. Rasmussen, and L. Frisk, Verification of ship collision frequency model, in Ship Collision Analysis. 1998: Balkema, Rotterdam. p. 117-121.
14. Ramboll, Navigational safety in the Sound between Denmark and Sweden (Öresund). 2006, Ramboll Danmark A/S: Denmark.
15. Gucma, L. The method of navigational risk assessment on waterways based on generalised real time simulation data. in International conference on marine simulation and ship maneuverability. 2006. Terschelling, Netherlands.
16. Kaneko, F. A method for estimation of grounding frequency by using trajectories of ships and geometry of seabed. in International Conference on Collision and Grounding of Ships (ICCGS). 2010. Espoo, Finland.
17. Lehman, A., JMP for Basic Univariate and Multivariate Statistics : A Step-by-step Guide. 2005, Cary, NC, USA: SAS Press.
18. Seong, Y. C., et al., The relation with width of fairway and marine traffic flow. TransNav-International Journal of Marine Navigation and Safety of Sea Transportation 2012. 6(3): p. 317-321.
19. Kujala, P., et al., Analysis of the Marine Traffic Safety in the Gulf of Finland. Reliability Engineering and System Safety 2009. 94(8): p. 1349-1357.
20. Pearl, J., Causal inference in statistics: An overview, R-350, Editor. 2009, Computer Science Department, University of California: Los Angeles. p. 96-146.
21. Neufeld, E. and S. Kristtorn, Does non-correlation imply non-causation. International Journal of Approximate Reasoning, 2007. 46: p. 257-273.