A Method for Assessing a Causation Factor for a Geometrical MDTC Model for Ship-Ship Collision Probability Estimation

J. Montewka*, F. Goerlandt, H. Lammi & P. Kujala
*Aalto University, School of Engineering, Finland; Maritime University of Szczecin, Poland
Aalto University, School of Engineering, Finland

ABSTRACT: In this paper a comparative method for assessing a causation factor for a geometrical model for ship-ship collision probability estimation is introduced. The results obtained from the model are compared with the results of an analysis of near-collisions based on recorded AIS data and then with the historical data on maritime accidents in the Gulf of Finland.

The causation factor is obtained for three different meeting types, for a chosen location and prevailing traffic conditions there.

1 INTRODUCTION

The MDTC (Minimum Distance To Collision) model for ship-ship collision probability estimation is a geometrical model, with a detailed description given in the following papers: (Montewka et al. 2010), (Montewka et al. 2011). In order to provide the probability of an accident, the model uses the commonly adopted approach, which combines a frequency of ship-ship meeting situations given an assumption of blind navigation, and a causation factor, which quantifies the proportion of cases in which such a meeting ends up as a collision, due to human or technical reasons.

The causation factor is a sensitive part of a model, very much location dependent, thus it is not justified to use the same value for the different models (Gluver and Olsen 1998). Applying a causation probability value derived from a study in another sea area may save some effort, but then the actual conditions are not addressed at all (Hanninen and Kujala 2009).

Two approaches can be recognized in the literature in order to estimate the causation factor. The simplified approach is based on a historical data, where the causation factor is assumed a ratio between the registered number of accidents and the estimated number of collision candidates (Fujii and Siobara 1971), (MacDuff 1974), (Inoue and Kawase 2007).

A second approach is more sophisticated, based on the concept of either event tree (Pedersen 1995), (Martins and Maturana 2010) or Bayesian Networks (DNV 2003), (Hanninen and Kujala 2009). This way of modelling is undoubtedly more time consuming than the first approach, however it allows getting an insight into the chain of events leading to an accident instead of providing just a number.

In order to determine the causation factor for the MDTC model for three different ship-ship encountering types (crossing, head-on and overtaking), we based our study on a modified first approach, which is relatively quick and straightforward thus robust. We perform two stage analysis, which combines the statistical data on maritime accidents and an analysis of near-collisions based on recorded AIS data.

The causation factor is being defined here as a ratio between the modelled number of collision candidates and the actual number of accidents. However the available statistics on maritime accidents are not very detailed, and the type of an accident is not included there. Thus there is a need to find a proxy between a recorded number of accidents and a modelled number of collision candidates (Heinrich et al. 1980), (Inoue and Kawase 2007), (Gucma and Marcjanc 2010).
It seems justified to analyze the safety of navigation on the basis of the numbers of both accidents and near-miss situations. Such a combination of analyses may better reflect the collision hazard, as pointed out by (Inoue et al. 2004) and (Inoue and Kawase 2007).

In air transportation there has been a tendency to seek out proxy for aviation safety. One commonly used measure is that of the “air-miss”, often called a "near-miss". According to (Button and Drexler 2006) "a near-miss involves an aircraft intruding upon a predetermined safety zone or envelope around another aircraft". The reporting procedures of near-miss in aviation are well founded providing valuable statistics. In the maritime sector similar procedures are missing, thus the near-miss can be detected only by analysis of recorded data and back propagation of recorded events.

Following this idea, this paper proposes also a methodology to evaluate the occurrence of near ship-ship collisions in an open sea area, based on the AIS data. The method for near-collisions analysis presented in this work is rooted in a well-established concept of a ship domain proposed by (Fujii and Tanaka 1971). An overview of the near collision detection method is then given and applied to the summer traffic in the Gulf of Finland.

Finally, we compare the results obtained from the MDTC model, expressed as the number of "collision candidates" with the number of near-collisions and the number of accidents recorded in the chosen area of the Gulf of Finland. This approach allows us to quantify the number of modelled "collision candidates", with blind navigation assumption behind, to the number of cases that ended up as close encounters, where collision evasive actions were taken. Such quantification is carried out for three major types of meeting scenario (crossing, head-on, overtaking). By combining this accurate enough data with an average annual number of accidents that happened (which are random, and almost non-predictable), the causation factor for the MDTC model is obtained.

2 RESEARCH MODEL

2.1 Accident analysis

The annual number of ship-ship collisions in the analyzed location of the Gulf of Finland (the waterways junction between Helsinki and Tallinn) is obtained from HELCOM database, that covers a time period between 1987 and 2007 (Pettersson et al. 2010). During this time, three accidents of this type took place. Two of them happened during summer time, and one was related to the ice conditions, which are out of scope of the analysis presented in this paper.

According to the aforementioned statistics there was, on average, one summer collision per ten years. This assumption is simplified, as the rate of collision occurrence is random, as the first collision happened in 1996, second in 2001 and between the years 2001 and 2007 no summer collision happened in the area of investigation. Notwithstanding, we assume that the annual ship-ship collision frequency in the analyzed area equals 0.1.

Unfortunately, the database provided by HELCOM does not contain any information regarding type of ship-ship encounter, at which the accident took place. Thus it is not feasible to compare a modelled number of collision candidates in given encounter type (crossing, head-on, overtaking) with an appropriate number of the accidents. At this point the results of near-collisions analysis are utilized and considered a proxy between a model and the recorded accident data.

2.2 Near-collisions analysis

The near-collision analysis applied in this paper is based on a concept of a ship domain, which according to definition given by (Goodwin 1975), is the area around the vessel which the navigator would like to keep free of other vessels, for safety reasons.

Since the first introduction of the ship domain concept by (Fujii and Tanaka 1971), various researchers have attempted to quantify the size of this domain. An overview of the different proposed domains is given in (Wang et al. 2009). Even though the ship domain is a well-established concept, certain problems with the application can be identified as pointed out by (Jingsong et al. 1993). Domains can be classified by their shape: circular, elliptical and polygonal domains. A distinction can also be made between fuzzy domains and crisp domains. Fuzzy domains such as that proposed by (Pietrzykowski 2008) and (Wang 2010) seem preferable in terms of
safety analysis of marine traffic, but are at present still under development. Crisp domains use a simple classification of a situation between safe or unsafe, which evidently is a simplification. Moreover, the sizes of the domains proposed in the literature vary quite significantly (Wang et al. 2009).

In this paper, the smallest ship domain found in the literature, by (Fujii and Tanaka 1971), is applied. This is justifiable, since the aim of the method proposed in this paper is finding the most critical encounters between ships. This domain is defined as an ellipse with the major axis along the ship’s length (LOA) and the minor axis perpendicular to the ship’s beam, as illustrated in Figure 1. The half-length of the major axis is taken as 4LOA while the half-length of the minor axis is taken as 1.6LOA. A number of comments should be made in the use of this domain:

− the domain is symmetric, which implies that the possible influence of the COLREGs is not taken into account;
− another consequence of this symmetry is the fact that passing behind the stern is considered as dangerous as passing in front of the bow;
− in the meeting between ships, the largest ship has the largest domain; this means that for the largest vessel, the situation is classified as dangerous, whereas for the smallest vessel, the situation may still be evaluated as safe;
− the domain is affected by ship length only, neither ship type nor hydrometeorological conditions are included in the analysis.

However the latter can be supported by the recent research, which revealed that the ship domain has a relatively low correlation with the sea state and wind force (Kao et al. 2007).

In this section, a brief description of analysis of AIS data in order to estimate a number of near-collisions in the selected area of the Gulf of Finland is given. Recorded AIS data consists of millions of data points, containing static and dynamic information regarding a ship. In order to analyze the maritime traffic in the GOF, this data need to be grouped into routes. Routes are defined here as a set of trajectories between a departure and arrival harbor as introduced by (Goerlandt and Kujala 2011). The AIS data is first gathered per ship, based on the MMSI number. After sorting this data chronologically, the data per ship is further split up to form individual ship trajectories, using a methodology described by (Aarsther and Moan 2009). These trajectories are then further processed and grouped per route. The sample rate of these vessel positions in the trajectories is about 5 minutes on average. In order to enable a comparison between vessel positions at exactly the same time instant, the trajectory data is artificially enhanced to contain data for each second. The extrapolation for the vessel position is performed using an algorithm suitable for data in the WGS-84 reference frame following (Vincenty 1975). The ship speed is linearly interpolated between known values. It should also be noted that certain vessel types are not taken into account into the analysis, like tugs are left out of the analyzed database. This is done because these vessels are meant to operate in a close vicinity of merchant vessels. The near collision detection algorithm is shown in Figure 2.

The basic idea is to scan the database for events where the ship contour of one vessel (i.e. the ship area in terms of ship length and width) enters the ship domain of another vessel. If the domain is violated, the event is labeled as a near collision and relevant details such as time of occurrence, location, encounter type, ship types and ship flags are stored for further analysis. The near collision detection algorithm is encoded in MATLAB.

**Figure 2: Near collision detection algorithm**
The algorithm starts with evaluating whether or not the trajectories of the two considered vessels occur in an overlapping timeframe. If so, the closest distance between vessel positions for contemporary time instances is computed using an algorithm appropriate for geodetic computations according to (Vincenty 1975). If this closest distance between points in trajectories is smaller than the extreme value of the ship domain, the actual vessel contour in terms of length and width are constructed for the smaller vessel and the ship domain is constructed for the larger vessel, for each second. Concurrent ship domain and a vessel contour are evaluated to overlap or not. If there is an overlap of a ship domain, the relevant situational data is stored. If there is no overlap, the next case is investigated.

In the analysis of the locations of the near collisions, a distinction is made between three different encounter situations, as defined in the Collision Regulations by (Organization 2002). Thus crossing, head-on and overtaking are considered. Having the data for the whole Gulf of Finland, we focus on a selected area, which is a crossing of waterways between Helsinki and Tallinn. The area is bounded by the following meridians: 024.5deg E and 025deg E and the parallels: 59deg N - 60deg N.

The results obtained in the course of the analysis, for the time period analyzed (01.04.2007-30.10.2007) are depicted graphically in Figure 3.

In the yearly perspective the elliptical Fujii domain leads to 14 ship domain violations for head-on encounters and 95 for crossing encounters. However, 252 cases are identified for overtaking encounters. This is due to the fact that the Fujii domain does not take the regulation of traffic in terms of traffic separation schemes into account. In order to get more meaningful results, a heuristic solution for this is applied, by requiring that the number of domain violations for overtaking is equal to the average number of critical encounters for head-on and crossing (labeled "Overtaking adjusted" in Table 1). To this effect, the Fujii domain is evaluated with a reduced width of $1.25L_{max}$ for overtaking encounters (as opposed to the original $1.6L_{max}$), where $L_{max}$ is the length of the largest vessel in the encounter.

### Table 1: The annual number of near-collision events in the waterways crossing in the Gulf of Finland.

<table>
<thead>
<tr>
<th>Ships meeting</th>
<th>Annual number of near-collisions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossing</td>
<td>95.0</td>
</tr>
<tr>
<td>Head-on</td>
<td>14.0</td>
</tr>
<tr>
<td>Overtaking</td>
<td>252.0</td>
</tr>
<tr>
<td>Overtaking adjusted</td>
<td>54.5</td>
</tr>
</tbody>
</table>

In the yearly perspective the elliptical Fujii domain leads to 14 ship domain violations for head-on encounters and 95 for crossing encounters. However, 252 cases are identified for overtaking encounters. This is due to the fact that the Fujii domain does not take the regulation of traffic in terms of traffic separation schemes into account. In order to get more meaningful results, a heuristic solution for this is applied, by requiring that the number of domain violations for overtaking is equal to the average number of critical encounters for head-on and crossing (labeled "Overtaking adjusted" in Table 1). To this effect, the Fujii domain is evaluated with a reduced width of $1.25L_{max}$ for overtaking encounters (as opposed to the original $1.6L_{max}$), where $L_{max}$ is the length of the largest vessel in the encounter.

#### 2.3 Collision probability modelling

The MDTC model, which is a geometrical model, estimates a probability of collision between two ships based on a well founded formula (Kristiansen 2004):

$$P = N_A P_C$$

where $N_A$ is the number of collision candidates, often named a geometrical probability of a collision course and $P_C$ is the causation probability, also called the probability of failing to avoid a collision when on a collision course. A ship on a collision course is called a collision candidate, which may end up as a collision as a result of technical failure or human error. The causation probability quantifies the proportion of cases in which a collision candidate ends up as a collision.

As a number of collision candidates $N_A$ depends on a number of factors, which are described in the following part of this chapter, the input data should be carefully chosen and interpreted before an analysis is carried out. The input values are location dependent, and within a specific location they are very often also time dependent, for instance:

- an intensity of traffic in the given area (if scheduled traffic is observed over the given area, the intensity of ships will change in the course of the day),
- a frequency of occurrence of given ship type in the given area (in general it can be correlated with

![Figure 3: Results of the near collision analysis](image-url)
scheduled traffic, in certain hours more ships of
given type can be expected than in an- other time
spans).

It is also important to observe a correlation be-
tween ship’s main particulars and ship type if sto-
chastic modeling is adopted.

MDTC model applied in this study distinguishes
between three types of ships encounters, these are:
crossing, overtaking and head-on. The probability of
having an accident in case of vessels crossing each
other course, is calculated by means of the following
formula (Endoh 1982), (Montewka et al. 2010):

\[
N_{\text{cross}\,g} = \sum_{i,j} \frac{E[V_{ij}]\lambda_i\lambda_j}{V_iV_j\sin\alpha} \sin \alpha
\]

where \( E[V_{ij}] \) denotes the expected relative velocity
of all pairs of vessels of types \( i \) and \( j \), \( \lambda \) denotes
the intensity of the vessels of given type entering the
given waterway, \( V \) is the velocity of the vessels ac-
cording to type, and \( \alpha \) is the angle of intersection be-
tween the courses of vessels in groups \( i \) and \( j \).

In case of parallel meetings, namely overtaking
and head-on meetings, the common formula is used,
and the difference is in a value of intersection an-
gle \( \alpha \). In case of overtaking \( \alpha < 10\deg \) and in case of
reciprocal courses \( 175\deg < \alpha < 185\deg \).

\[
N_{\text{parallel}} = T_0P_0P_{\text{time}}
\]

where \( T_0 \) is the overtaking rate (the number of ves-
sels which will overtake another while on parallel
courses, irrespective of the passing distance), \( P_0 \) is
the probability that the vessels come close to each
other and \( P_{\text{time}} \) is the probability that these two ships
being close to each other will meet in a certain time
period. The latter also reduces the theoretical possi-
bility of ship colliding themselves and is estimated
for scheduled traffic between Helsinki and Tallinn.
This probability is not taken into account in case of
E-W traffic, which is more random in nature. The
overtaking rate is obtained by means of the follow-
ing equation (Endoh 1982), (Montewka et al. 2010):

\[
T_0 = \frac{N^2}{2L} E[V_{ij}]
\]

where \( N \) is the expected number of vessels in the
waterway on parallel courses, \( L \) is the length of wa-
terway, and \( E[V_{ij}] \) denotes the expected relative ve-
locity of all pairs of vessels of types \( i \) and \( j \). The ex-
pected relative velocity between two vessels is
determined as follows:

\[
E[V_{ij}] = \sqrt{V_i^2 + V_j^2 - 2V_iV_j\cos\alpha}
\]

where \( V_i \) is the velocity of a vessel of given group, \( \alpha \)
means the angle of intersection, which is defined as
the difference between the courses of vessels in
groups \( i \) and \( j \).

The probability that the vessels come to a dis-
tance, that results in a collision \( (P_0) \) is simply esti-
mated as follows:

\[
P_0 = P\left( d < \frac{B_i + B_j}{2} \right)
\]

where \( d \) is the distance between two ships while
overtaking and \( B \) is the breadth of a vessel of a given
class \( i \) and \( j \). In order to obtain the results as close as
possible to the results of near-collisions analysis, the
same criteria have to be used. Thus the critical dis-
tance for ships on parallel courses is adopted from
the near-collisions algorithm, and equals 1.25LOA.
Figure 5: Intensities of marine traffic streams

Figure 6: Histograms of main particulars of ships over analyzed area.
Maritime traffic in the area is assumed to be a stochastic process, and is modelled by means of random sampling and Monte-Carlo methodology. The initial traffic database is decomposed into four smaller databases, according to the four main traffic streams (Figure 4). Then each stream is modelled separately, taking into account the non uniform distribution of ships in time over each stream. The histograms of parameters used in the course of marine traffic analysis are presented in Figures 5 and 6.

The MDTC value, which acts as an input for the equation 2, is drawn from the appropriate chart (Figures: 7, 8, 9). The charts were obtained in the course of an analysis with the use of a model of ship motion given the maneuvering pattern and a ship type (Montewka et al. 2011). The maneuvering pattern, which decides if both of the ships involved in collision situation perform collision evasive actions or only one of them, is chosen randomly with the same probability of occurrence for each of them ($p = 0.5$). Such an assumption may sound simplified, however there is not enough evidence in the literature to disregard it. In case where the maneuvering pattern one is chosen, the algorithm checks if there is a tanker involved, if so then an appropriate MDTC value only for tankers is chosen.

![Figure 7: The obtained MDTC chart for the maneuvering pattern No 1 (Montewka et al. 2011)](image1)

![Figure 8: The obtained MDTC chart for tankers - the maneuvering pattern No 1 (Montewka et al. 2011)](image2)

The annual number of collision candidates obtained with the use of MDTC model is presented in Table 2. The results are divided into three meeting scenarios (crossing, head-on and overtaking). Within these scenarios there are different sub-scenarios which represents meetings of ships sailing in various streams.

<table>
<thead>
<tr>
<th>Ships meeting</th>
<th>Annual number of collision candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossing</td>
<td>5538</td>
</tr>
<tr>
<td>Head-on N-S</td>
<td>1.0</td>
</tr>
<tr>
<td>Head-on E-W</td>
<td>57.0</td>
</tr>
<tr>
<td>Head-on All</td>
<td>58.0</td>
</tr>
<tr>
<td>Overtaking N-N</td>
<td>164</td>
</tr>
<tr>
<td>Overtaking S-S</td>
<td>156</td>
</tr>
<tr>
<td>Overtaking E-E</td>
<td>28</td>
</tr>
<tr>
<td>Overtaking W-W</td>
<td>34</td>
</tr>
<tr>
<td>Overtaking All</td>
<td>382</td>
</tr>
</tbody>
</table>

### 3 RESULTS

In the course of presented analyses we obtain a data regarding near-collisions, number of accidents and modelled number of collision candidates for the specific location in the Gulf of Finland.

The aim of this research is to develop a causation factor for the MDTC model by means of a comparative study. The values of the causation factor ($P_C$) are strongly location dependent, as the original studies regarding this parameters have been conducted in the specific locations (eg. straits in Japan, the Dover Strait) it is difficult to assess how the results obtained there can be transferable to other sea areas. The $P_C$ value is also highly dependent on a geometrical model used for the probability of ship accident estimation, thus transferring the same value between different models seems not justified from the scientific point of view.
In our approach we estimate the causation factor that is related to the MDTC model, based on the following formula:

\[ P_c(m) = \text{SHF}_m \sum_m \frac{N_d}{N_{\text{near-coll}}} \]  

(7)

\[ \text{SHF}_m = \frac{N_{\text{near-coll}}(m)}{N_{\text{coll-cand}}(m)} \]  

(8)

where \( \text{SHF} \) is a ship handling factor defined for each type of meeting \( m \) individually ( \( m = \text{[head-on, overtaking, crossing]} \)), for the specific value of \( \text{SHF} \) see Table 3, \( N_{\text{near-coll}} \) is a number of observed near-collisions, \( N_{\text{coll-cand}} \) is a number of modelled collision candidates and \( N_d \) is a number of recorded accidents. As a result the causation factors for three types of ship/ship encounter were estimated (Table 4).

Ship handling factors presented in Table 3 govern a ship handling process, showing a difference between blind navigation model and real traffic for different encounter types.

Table 3: The ship handling factor (SHF) for three types of ship/ship encounter, for a specific location in the Gulf of Finland.

<table>
<thead>
<tr>
<th>Type of meeting</th>
<th>The SHF for given events</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossing</td>
<td>1.7 * 10e - 2</td>
</tr>
<tr>
<td>Head-on</td>
<td>2.4 * 10e - 1</td>
</tr>
<tr>
<td>Overtaking</td>
<td>1.4 * 10e - 1</td>
</tr>
</tbody>
</table>

In Table 4 the values of the causation factor for the MDTC model are gathered. However further research which would cover the greater sea area leading towards a better definition of the causation factor should be carried out.

The numbers for the causation factor proposed here consider a specific geometrical model (MDTC), ordered traffic with waterways crossing, continous surveillance from VTS stations, presence of Traffic Separation Schemes and an intense RoPax cross traffic. The proposed causation factors make a distinction between type of ship-ship encounter. The model is applicable only for the "summer traffic", which means, that presence of ice is not considered.

Table 4: The causation factor for the MDTC model for three types of ship-ship encounter.

<table>
<thead>
<tr>
<th>Type of meeting</th>
<th>The causation factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossing</td>
<td>1.04 * 10e - 5</td>
</tr>
<tr>
<td>Head-on</td>
<td>1.46 * 10e - 4</td>
</tr>
<tr>
<td>Overtaking</td>
<td>0.85 * 10e - 4</td>
</tr>
</tbody>
</table>

The general relations between analyzed types of event (modelled number of collision candidates, observed number of near-collisions and recorded number of accidents) for the analyzed location are depicted in Figure 10.

4 CONCLUSIONS

This paper addresses a problem of defining the causation factor for a given geometrical model. We propose a straightforward methodology, which is based on recorded near-collisions (obtained in the course of AIS data analysis) and actual collisions (obtained from HELCOM accidents database). The method establishes the ratios between the recorded number of accidents, the recorded number of the near-collisions and the modelled number of the collision candidates. Knowing these values, it is possible to define a causation factor that constitutes a link between a geometric model for ship-ship collision frequency estimation and a number of accidents due to the given parameters of marine traffic and surroundings.

Making a comparative study we defined the causation factors for the MDTC model, for three ship-ship encounter types. The estimated values of causation factors for the selected area of the Gulf of Finland and given types of vessels sailing there are of the following orders of magnitude: \( 10e - 5 \) for ships crossing and \( 10e - 4 \) for ships meeting each other on parallel courses.

Although the methodology behind this analysis is straightforward, the results are promising, however there is a need for more extensive analysis, that would cover a larger sea area.

ACKNOWLEDGMENT

The authors appreciate the financial contributions of the following entities: the EU, Baltic Sea Region (this research was founded by EfficienSea project), Merenkulun säätiö from Helsinki, the city of Kotka and the Finnish Ministry of Employment and the Economy.
REFERENCES


