Probability modelling of vessel collisions

Jakub Montewka *, Tomasz Hinz, Pentti Kujala, Jerzy Matusiak

Aalto University, School of Science and Technology, Department of Applied Mechanics, Marine Technology, P.O. Box 15300, FI-00076 AALTO, Espoo, Finland

A R T I C L E   I N F O

Article history:
Received 9 September 2009
Received in revised form 16 January 2010
Accepted 18 January 2010
Available online 25 January 2010

Keywords:
Marine traffic safety
Collision probability
Minimum distance to collision (MDTC)
Ship motion model
Gulf of Finland

A B S T R A C T

Among engineers, risk is defined as a product of probability of the occurrence of an undesired event and the expected consequences in terms of human, economic, and environmental loss. These two components are equally important; therefore, the appropriate estimation of these values is a matter of great significance. This paper deals with one of these two components—the assessment of the probability of vessels colliding, presenting a new approach for the geometrical probability of collision estimation on the basis of maritime and aviation experience. The geometrical model that is being introduced in this paper takes into account registered vessel traffic data and generalised vessel dynamics and uses advanced statistical and optimisation methods (Monte Carlo and genetic algorithms). The results obtained from the model are compared with registered data for maritime traffic in the Gulf of Finland and a good agreement is found.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

At present the most common approach used for static collision probability assessment is rooted in the research carried in the 1970s by Fujii et al. [1,2] and MacDuff [3]. According to this, the probability of a collision is defined as follows:

\[ P = N_g P_c \]

where \( N_g \) is the geometrical probability of a collision course and \( P_c \) 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. The value of the causation probability for this analysis is adopted from a state-of-the-art model based on a Bayesian Belief Network developed in earlier research [4–6]. This paper focuses on modelling the geometrical probability of a collision, and presents a novel approach that takes into account ship manoeuvrability.

The Fuji’s model, which has been commonly used in recent decades, won the popularity among researchers mainly due to its simplicity and robustness. However it has also some drawbacks, it lacks ship dynamics and it assumes that a collision happens when two vessels come to a distance defined as the “collision diameter” \( (D_{ij}) \) depicted in Fig. 1, which means almost physical contact. Such an assumption may lead to an understanding that in any encounter between two vessels at a distance greater than \( D_{ij} \) they are able to avoid a collision, which in most cases is not true. Therefore one may suspect that the model may not fully reflect the real interaction between vessels at close quarters. Despite the drawbacks the model was adopted by Pedersen, and with minor modifications was used to determine the safety of navigation in many European waters [7–9]. Hence in Europe it is mostly known as the Pedersen model.

Another approach based on a ship domain theory was applied to perform a risk analysis for a large suspension bridge by Pedersen [10], in which the critical distance between two ships was assumed a constant value, and only head-on encounters were considered. A similar approach was used by Fowler and Sørgård [11] in their model, which assumed a critical situation as a situation when two vessels come to close quarters, crossing within 0.5 Nm of each other, which was fixed regardless of the crossing angle. A series of papers utilizing the ship domain approach for ship-fixed object collision assessment was published also by Gluver and Olsen [12].

A model for encounter probability estimation proposed by Kaneko [13] defines a critical area of an optional form of a closed boundary, around a ship, whose violation means collision. Kaneko in his model recognizes two shapes of the critical area: rectangular and circular, but again the size of the area is fixed, and ship manoeuvrability is not considered.

In the last decade two trends became popular in the field of marine transportation safety assessment; one in which methods based on Markov processes are utilized [14,15], and another that makes use of the Bayesian Belief Network [16–18]. However ship dynamics in terms of manoeuvring abilities is rarely mentioned there as well.
At present it is very popular to apply models based on marine traffic simulation, which are commonly known as fast time numerical navigators (FNNs), in the assessment of collision and grounding frequencies [19–23]. But whenever possible the geometric approach is implemented to the FNN, simply because of the speed and stability of the former, and similar to the previous approaches it is a very common practise to consider the critical distance between two ships as a constant value [22]. Although a definition of the critical distance between two ships being on a collision course is often met in papers dedicated to collision avoidance systems [24–26], it is very rarely transferred directly to collision and grounding assessment process.

In the approach presented in this paper, it is assumed that a collision between two vessels becomes reality when the distance between the vessels is not enough to perform efficient anti-collision manoeuvres. The space and time required for a given vessel to perform a given manoeuvre depends mostly on her hydrodynamics and manoeuvrability features. For the first time the idea of ship's manoeuvrability implementation into a collision assessment model was presented by Curtis [27], but his model was limited to one ship type, which was a very large crude carrier (VLCC) and only overtaking and head-on situations were considered.

The main intention of this paper is to obtain the minimum distance between two vessels on collision courses that allows them to perform efficient anti-collision manoeuvres, using a ship in-plane motion model. It is assumed that these two vessels are navigating in deep water, they start manoeuvres simultaneously, and the rudder and engine settings are constant during the time of the manoeuvres.

In the following chapters this critical distance is called the minimum distance to collision (MDTC) and it is an input value for the model that assesses the probability of a collision. The experiment was conducted for 4 vessel types of different sizes, proceeding at their maximum velocities. This finally produced 8 meeting scenarios. Each scenario was repeated for 17 crossing angles.

The model is expected to work with data transmitted by a ship-borne automatic identification system (AIS) as an input. Therefore it has been tailored to be used with limited input values. The model of ship dynamics that is implemented uses approximations that allow an AIS message and information about the ship type to be sufficient to give some realistic output.

The analysis presented in the paper concerns the most dense sea area in the Gulf of Finland (GoF), in terms of marine traffic, which is a junction of waterways between Helsinki and Tallinn.

2. Collision probability modelling

The collision probability prediction model presented in this paper is based on a molecular collision model and is supported by a model of ship dynamics. The molecular collision model is one of the several models used in aviation for the probability analysis of aircraft colliding [28]. Its application for marine purposes was presented by MacDuff [3], but in his approach ship manoeuvrability was not considered. This paper presents a two-dimensional model, reduced from the original three-dimensional case for the purposes of marine navigation. In this model the vessel is represented as a particle surrounded by a disc of a given radius (Fig. 2a), which constitutes a “no-go area” for other objects. Some researchers define the vessel static domain in the similar way [26,29], while others call it the bumper [30] or the guarding ring [31]. The definition of the domain expresses it as the area around the vessel that the navigator wants to keep clear of other vessels or objects [32,33], and therefore a violation of the vessel’s domain is not tantamount to a collision. In the presented model a violation of the outer boundary of the disc is equal to a collision (Fig. 2). A collision between two vessels (two discs) is described as an overlap of these two discs. The occurrence of such an overlap is equivalent to an event in which a point representing the centre of one vessel enters the disc of which the radius equals the sum of the radii of two original discs (Fig. 2b).

The diameter of the greater disc (Fig. 2b) is not fixed, and is computed for each type of vessel and encounter individually, thus it changes with situations. The value of that disc diameter is considered the MDTC. If the distance between these two vessels becomes less then MDTC, it means that a collision cannot be avoided by any manoeuvres, and the vessels will collide. In other words, as long as the disc of radius MDTC is not violated by another vessel no collision will take place. This situation and relations between MDTC, safe passing, and collision are presented graphically in Fig. 3. The main factors relevant to determine the MDTC value are: the vessels’ manoeuvrability, the angle of intersection (labelled $\alpha$ in Fig. 3a), and the relative bearing from one vessel to the other (angle labelled $\beta$ in Fig. 3a). Recent research has revealed that the vessel’s safety domain, which may be comparable to MDTC, had a relatively low correlation with the sea state and wind force [31]. Therefore the hydro-meteorological conditions are not considered.

2.1. Model formulae

This chapter contains the formulae used in the model, and presents results of the analysis of the three main types of vessel encounters: overtaking, head-on, and crossing.

Overtaking means a situation in which two vessels are proceeding on the same route, lying on almost parallel courses, with courses difference not exceeding $10^\circ$.

Head-on means a situation in which vessels are lying on almost reciprocal courses, and the course difference falls in the range $180 \pm 10^\circ$.

Crossing means a situation in which the difference between two vessels’ courses falls in the range $10–170^\circ$.

![Fig. 2. Representation of vessels as discs and definition of collision situation.](image-url)
2.1. Vessels on parallel courses — overtaking

In this case, the vessels are navigating along the same route, on mutually parallel courses, with the difference between their courses not exceeding 10 degrees, but at different velocities. The geometrical probability of a collision between two vessels while overtaking also referred to as the numbers of candidates for collision during overtaking, is expressed as follows:

\[ N_{\text{overtaking}} = T_0 P_0 \]  \hspace{1cm} (2)

where \( T_0 \) is the overtaking rate and \( P_0 \) the probability that the vessels’ domains are violated during overtaking. The overtaking rate is the number of vessels that will overtake another while on parallel courses, irrespective of the passing distance. The overtaking rate is not a collision frequency, unless the waterway width is zero. It is calculated with the following formula [28]:

\[ T_0 = \frac{N^2}{2L} E[V_i] \]  \hspace{1cm} (3)

where \( N \) is the expected number of vessels in the waterway on parallel courses, \( L \) the length of waterway, and \( E[V_i] \) denotes the expected relative velocity of all pairs of vessels of types \( i \) and \( j \). The expected relative velocity is determined as follows [28]:

\[ E[V_i] = \frac{1}{N - 1} \sum_{j \neq i} E[V_{ij}] \]  \hspace{1cm} (4)

\[ E[V_{ij}] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sqrt{(V_x^2 + V_y^2 - 2V_i V_j \cos \theta)} P_i(V_i) P_j(V_j) \, dv_i \, dv_j \]  \hspace{1cm} (5)

where \( V_i \) is the velocity of a vessel of given group [m/s], \( \theta \) the angle of intersection, which is defined as the difference between the courses of vessels in groups \( i \) and \( j \), and \( P_i(V) \) denotes the probability density function of velocity for a given type of vessel. The probability that the vessel domains will be violated while overtaking \( (P_0) \) equals the probability of the event that two vessels will pass each other at a distance less than the adopted value and is expressed as follows:

\[ P_0 = P(\sin \alpha < \frac{B_1 + B_2}{2}) \]  \hspace{1cm} (6)

where \( \alpha \) is the distance between two ships while overtaking and \( B \) the breadth of a vessel of a given class.

2.1.2. Vessels on parallel courses — head-on

The number of collision candidates during head-on meetings is calculated using the formulae presented in Eqs. (2)–(6). It is assumed that the differences between the two vessel courses are to be within \( 180 \pm 10^\circ \) to consider such a situation as a head-on meeting.

2.1.3. Vessels on crossing courses

To calculate the collision rate at the intersections of waterways, it is assumed that the vessels are entering the waterway with a given velocity, which is modelled by a continuous distribution, and with a given intensity, which is assumed to be constant. The processes of the flow of vessels into waterways are independent. The number of collision candidates is determined on the basis of the following equation [28]:

\[ N_{\text{crossing}} = \sum_{ij} \frac{RE[V_i] \lambda e^{-l}}{\sqrt{V_i V_j} \sin \alpha} \]  \hspace{1cm} (7)

where \( R \) is the MDTC value, \( \lambda \) denotes the intensity of the vessels entering the waterway, \( V \) is the velocity of the vessels according to type, and \( \alpha \) the angle of intersection of the waterways. The number of collision candidates depends on the vessels’ velocities and dimensions, the traffic intensity, and the angle of intersection of the waterways. The rate does not depend on the lateral distribution of vessels across the waterway, whereas it is crucial in the case of overtaking and head-on situations. Eqs. (6) and (7) are based on different assumptions and are intended for different types of meeting. Eq. (7) is used in the case of crossing cases and includes a MDTC value, which Eq. (6) does not. On the other hand the Eq. (6) is used in the case of overtaking and head-on meetings and considers a spatial distribution of vessels across waterways, which is not relevant in Eq. (7) as mentioned above. It should be understood that these two equations are not to be used interchangeably, they are independent. There is a strict range of angles they may be used within, as was defined at the beginning of this chapter. Formulae used beyond their angles’ range may produce wrong results.

2.2. Assessment of minimum distance to collision

To determine the value of MDTC and the factors that may affect it, an experiment using the hydrodynamic model of ship motion was conducted and several crossing-type meeting scenarios were simulated. Analysis was carried out for the following:

\begin{itemize}
  \item 3 types of vessels (container carrier, passenger vessel, and tanker);
  \item 9 meeting scenarios, described in the next part of the paper;
  \item 17 crossing angles varying from 10° (almost overtaking) to 170° (almost head-on), with 10° increments (Fig. 4);
\end{itemize}
4 types of manoeuvres conducted by both vessels to avoid collision, as presented in Table 1.

The following assumptions were made:

- initially the vessels are proceeding at full sea speed;
- the vessels are fully laden;
- the vessels start their manoeuvres simultaneously;
- the settings of the steering gear and propulsion during the manoeuvre are constant;
- the initial course of vessel A is always 360° and vessel B changes her initial course in each consecutive trial by 10°;
- the initial relative bearing from vessel A to vessel B is 45° starboard;
- only the “turning away manoeuvres” performed by means of rudder are considered;
- the influence of weather conditions is omitted;
- the curvature of the Earth is omitted because of the small distances being considered.

The “turning-away manoeuvre” implies course alteration away from each other to avoid collision and to shorten the time at close quarters. The data presented in Table 1 define the turning-away manoeuvres according to the initial course of vessel B.

The initial relative positions of the vessels are presented in Fig. 4. The crossing angle number depends on vessel B’s course. Number 1 represents vessel B on a course of 350° and number 17 describes a meeting where vessel B’s course equals 190°.

The most intriguing question is if and how the types of meeting vessels affect the value of MDTC. For this purpose the following meeting scenarios were simulated, appropriate nomenclature was adopted, then used in Fig. 7:

- two container ships of the same size (cont_cont),
- two container ships of different sizes (cont_diff),
- two tankers of the same size (tanker_tanker),
- two tankers of different sizes (tankers_diff),
- passenger vessel–tanker (tanker_pass),
- two passenger vessels (pass_pass),
- passenger vessel–container ship (pass_cont),
- ro-ro–container ship (ro-ro_cont),
- two ro-ro carriers (ro-ro_ro-ro).

The analysis was conducted for the traffic in the Gulf of Finland (GoF); therefore the sizes of typical “GoF vessels” are considered. A detailed analysis of marine traffic in this area is conducted in Section 3, and typical “GoF vessels” are presented. The algorithm for MDTC assessment is presented in Figs. 5 and 6. The optimisation process whose aim is to obtain the minimum distance between the initial positions of two manoeuvring vessels (MDTC) is performed with the use of genetic algorithms. The aim of the optimisation process was to find the value of B0, which is the initial position of the centre of gravity of vessel B at time instant 0, as presented in Fig. 5. An iteration algorithm was used to determine the B0 value for a certain meeting scenario, by

Table 1

<table>
<thead>
<tr>
<th>Type of manoeuvre</th>
<th>Vsl_A to port, Vsl_B to stbd</th>
<th>Both to port</th>
<th>Both to stbd</th>
<th>Vsl_A to stbd, Vsl_B to port</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 190°–270°</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>&lt; 270°–350°</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Fig. 4. Vessel relative positions with three chosen crossing angles, before they start to manoeuvre.
finding a minimum value of $e_i$ as follows:

$$\min(e_i) = B_0$$

$e_i = B_i - A_i$

$0.6 \text{LOA} > \min(e_i) > 0.5 \text{LOA}$

(8)

where $B_i$ is the position of the centre of gravity of vessel B at the time instant $t_i$: $X_B, Y_B, t_i$. $A_i$ the position of the vessel A's centre of gravity at the time instant $t_i$: $X_A, Y_A, t_i$, $e_i$ the distance between the centres of gravity of two vessels at the time instant $t_i$ and LOA denotes the average value of length over all of two vessels involved. It was assumed that a distance between the centres of gravity of two appropriate vessels that is larger than 0.5 LOA implies non-contact meeting. The idea of the above process is presented graphically in Fig. 5, and values of $e_i$ for consecutive time instants are presented.

### 2.2.1. Results of numerical experiment of MDTC assessment

The results of the experiment are presented in Fig. 7, together with the corresponding “collision diameter” values (CD) used in Pedersen’s model [8]. The collision diameter and MDTC may be considered equivalent in principle, and therefore a comparison of results is performed, to show the discrepancies between these two values. Fig. 7 consists of a number of curves, which represent the meetings of different types of vessels. According to the legend
the values of MDTC and CD are non-dimensional, expressed in terms of the average length of the vessels involved. Depending on the angle of intersection, the discrepancies vary. For angles in the range 10–30° and 140–170° the values of MDTC and CD do not differ much but for the angles 90–110° the highest discrepancy is observed. This range of crossing angles is the most unfavourable in the sense of the value of course alteration required to avoid a collision.

Depending on the types of vessels engaged in the meeting, there are slight differences in the MDTC values. To prove whether these differences are statistically significant, an appropriate analysis was performed. A hypothesis was formulated that each MDTC curve was drawn from the same population, and therefore the null hypothesis cannot be rejected. Therefore the results allow it to be assumed that all MDTC values are drawn from the same population. Because of the limited survey sample, further analysis was based on the Monte Carlo simulation and both mean and standard deviation values were obtained. Afterwards, both MDTC and CD values at a 95% confidence level were calculated, and the results are presented in Fig. 8.

### 2.3. Ship dynamics modelling

A vessel’s manoeuvrability is estimated with the quasi-linear modular hydrodynamic model of the vessel in-plane motion. Ship motion is calculated from the following generalised formulae:

\[
\begin{align*}
\frac{du}{dt} &= \frac{(X_{\text{resist}} + X_{\text{prop}} + mrv)}{(m - X_u)} \\
\frac{dv}{dt} &= \frac{(-mru + Y_u + V_r + V_f + L)}{(m - Y_v)} \\
\frac{dr}{dt} &= \frac{(N_v + N_r + N_f + L_p/0.5)}{(I_{zz} - N_r)}
\end{align*}
\]

where \(u\) is the linear velocity along \(X\)-axis, \(X_{\text{resist}}\) the ship resistance (Eq. (10)), \(X_{\text{prop}}\) the propeller thrust (Eq. (11)), \(m\) denotes the ship mass, \(r\) is the angular velocity along axis \(Z\)-axis, \(v\)
the linear velocity along Y-axis, \( L \) the lift forces on the rudder (Eq. (12)), \( L_{pp} \) the length of the ship between perpendiculars, \( I_{xx} \) is a mass moment of inertia, and \( X_0, Y_0, Y_r, Y_t, N_k, N_l, N_u, N_r \) are hydrodynamics derivatives and are calculated using formulae derived from the literature [34–36].

\[
X_{\text{rest}} = -0.5 \rho u^2 SC_T \frac{1}{(1-t)}
\]

(10)

where \( \rho \) is the water density [tonne/m^3], \( S \) denotes the wetted surface of the ship hull (Eq. (18)), \( C_T \) the total resistance coefficient (Eq. (17)), and \( t \) is the thrust deduction factor (Eq. (16)). The propeller thrust is calculated with the following conditional equations:

\[
X_{\text{prop}} = \begin{cases} \frac{T}{T_{BP}} & T_{BP} < T \\ 1 & n > 0 \\ T > T_{BP} \end{cases}
\]

(11)

where \( T \) is the propeller thrust calculated by means of Eqs. (19) and (20), and \( T_{BP} \) denotes the bollard pull, obtained with the use of Eq. (21). The lift force on the rudder is calculated by means of the following formula:

\[
L = \frac{0.5 \rho u V_R^2 \pi \Lambda (A+1)}{(A+2)^2}
\]

(12)

where \( \rho \) is the water density [t/m^3], \( A_k \) the rudder area [m^2], \( V_R \) the velocity near the rudder [m/s], and \( \Lambda \) the aspect ratio calculated as follows:

\[
\Lambda = \frac{b^2}{\pi} A_k
\]

(13)

where \( b \) is the rudder length [m]. The rudder area is estimated on the basis of the following formula [36]:

\[
A_k = 0.01 L_{pp} T_{\text{max}} \left[ 1 + 50 C_f^2 \left( \frac{B}{L_{pp}} \right)^3 \right]
\]

(14)

where \( T_{\text{max}} \) is the maximum draught of the vessel [m], \( C_f \) denotes the block coefficient, and \( B \) the ship breadth [m]. The block coefficient is calculated using the following formula by Lee et al. [37]:

\[
C_B = 2.11 - \frac{4.7}{4.03 - \sqrt{L_{OA}}}
\]

(15)

where \( u \) is the linear velocity of ship [m/s] and \( L_{OA} \) the overall ship length [m]. The thrust deduction factor is estimated as follows [38]:

\[
t = 0.905 \left[ 0.17 + 0.2 C_f - 0.015 \left( \frac{L_{OA}}{B} \right) + 0.01 \left( \frac{B}{T_{\text{max}}} - 2.5 \right) \right]
\]

(16)

where \( T \) is the ship draught [m]. The total resistance coefficient is calculated using the following formula [39]:

\[
C_T = 8 / \pi \frac{K_f(j)}{J}
\]

(17)

where \( K_f(j) \) is a non-dimensional thrust coefficient and \( J \) an advance coefficient calculated with the use of Eq. (20). The wetted surface of the hull is calculated by means of the following formula [40]:

\[
S = \sqrt{\frac{B}{C_T - 0.2 (C_T - 0.55)}} + \frac{(B/T)}{1.7}
\]

(18)

where \( V \) stands for the volume of ship displacement [m^3]. The model of propeller thrust was calculated using the following formulae:

\[
T = K_f(j)p n^2 D^4
\]

(19)

\[
J = \frac{V_A}{nD}
\]

(20)

where \( n \) stands for propeller revolutions [min^{-1}], \( D \) is the propeller diameter [m], and \( V_A \) is the speed of advance of the propeller [m/s]. The values of \( K_f(j) \) are derived from the literature [41] for a screw of the Wageningen B series. In some cases, especially for low ship speed, Eq. (19) may give unrealistic results. To prevent these abnormally high thrust values, a limiting value was established, which corresponds to the bollard pull value. The following formulae are used to calculate the bollard pull, depending on the direction of propeller revolution:

\[
T_{BP} = \begin{cases} \frac{7.8 (P_D D)^{2/3}}{n} & \text{for } n > 0 \\ \frac{5.5 (P_D D)^{2/3}}{n} & \text{for } n < 0 \end{cases}
\]

(21)

where \( P_D \) is the power of the ship’s main engine [kW], which is estimated by means of Eq. (23). The propeller diameter is estimated with the use of the following formula:

\[
D = 0.6 T_{\text{max}}
\]

(22)

The vessel’s power is estimated on the basis of modified Admiralty formula:

\[
P_D = \frac{V^a \eta_{PD} C}{\eta_D C}
\]

(23)

where \( V \) is the vessel velocity [kn], \( \eta_{PD} \) the propulsion efficiency, \( C \) denotes the Admiralty constant, and \( a \) and \( b \) are exponents estimated for each type of vessel individually.

Each class of analysed vessels that was analysed was studied to find values of \( \eta_{PD} \) and \( C \) and exponents of \( V \) and \( V \). The research was focused on the four main groups of ships that operate most frequently across the Gulf of Finland: these are tankers, ro-ro, passenger vessels, and container carriers. For each class of vessel the following data were extracted from a ship database: length, breadth, draught, service speed, displacement, and main engine power. For the purposes of this paper, two ship databases were studied. One database was provided online by the Japanese ship classification society Nippon Kajii Kyokai, from which data concerning tankers, container carriers, and ro-ro vessels were extracted. Another source of particulars on vessels was the bulletins issued by the operators of passenger vessels that cruise across Gulf of Finland. For each type of vessel analysed, a sample of 60 vessels was extracted from the databases. The exception was the group of passenger vessels (20 vessels); the sample size was limited by the actual number of vessels in operation. Afterwards the optimisation method based on genetic algorithms was used to analyse the samples and to find the values of \( \eta_{PD} \), \( C \), and exponents of \( V \) and \( V \). The aim of the optimisation process was to minimise the error function (\( \varepsilon \)) and maximise the prediction interval (\( \Omega \)) simultaneously. These two are expressed as follows:

\[
\varepsilon = 100 \left( \frac{|P - P_D|}{P} \right) \% \%
\]

(24)

\[
\Omega = 100 - 100 \left( \frac{|N_P - N_{PD}|}{N_P} \right) \% \%
\]

(25)

where \( N_P - \varepsilon \leq N_{PD} \leq N_P + \varepsilon \)

(26)

where \( P_D \) denotes the predicted power of the ship’s main engine, \( P \) the power of the sample vessel, taken from the ship database, \( N_P \) the number of sample vessels for which the engine power is known (taken from ship database), \( N_{PD} \) the number of vessels for which the predicted power is within the adopted limit, and \( \varepsilon \) is the error function defined by Eq. (24). The results obtained are presented in Table 3.

2.4. Validation of applied ship dynamics model

Although the ship dynamics model used in this study is well known and has been validated many times [35,42], it is justified
to compare the results coming from the model with the results of sea trials to assess to what extent the simplification and approximation adopted in the present study influence the accuracy of the ship motion model. Validation was carried out for three ship types, the data for which are presented in Table 4.

In the experiment described in the previous section, whose aim was to determine the MDTC values, the evasive manoeuvre was based on the turning circle performed by both vessels. The validation of simulated manoeuvres was carried out in the domain of time and the time period investigated was 200 s. In Fig. 9 the histogram and maximum and minimum values of time required by vessels to evade a collision situation by performing a turning circle are presented. These values were computed in the MDTC determination process. The time was counted from the beginning of the taking of the anti-collision action until the minimum distance between these vessels min(\(e_i\)) was reached (Fig. 5). According to these data the maximum time required for vessels to perform a collision avoidance manoeuvre is 130 s.

### Table 4
Main particulars of ships for which the model was validated.

<table>
<thead>
<tr>
<th>Vessel type</th>
<th>L [m]</th>
<th>B [m]</th>
<th>T [m]</th>
<th>DWT [t]</th>
<th>Power [kW]</th>
<th>Speed [kn]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ro-ro</td>
<td>199.4</td>
<td>23.4</td>
<td>6.08</td>
<td>7400</td>
<td>18,900</td>
<td>20.9</td>
</tr>
<tr>
<td>Tanker</td>
<td>183.0</td>
<td>32.2</td>
<td>11.0</td>
<td>40,000</td>
<td>8848</td>
<td>13.7</td>
</tr>
<tr>
<td>Container carriers</td>
<td>220.0</td>
<td>32.24</td>
<td>10.5</td>
<td>42,500</td>
<td>26,270</td>
<td>22.2</td>
</tr>
</tbody>
</table>

Fig. 9. Time to closest distance between two vessels (\(e_i, \text{min}\)) performing collision avoidance manoeuvres.

Fig. 10. Estimated value of ship speed and course versus observed data.
In the validation process two parameters were compared: the ship's course and her speed during circle turning. Data from the sea trials of newly built ships were chosen as references. The results are presented in Fig. 10.

The graphs in Fig. 10 show good agreement between modelled values and the observed ones for vessels equipped with conventional single-propeller engines. Differences in the vessels' course and speed occur for vessels equipped with twin-screws engines (ro-ro). However, the differences that occur until the 100 s time instant are acceptable in all cases. Time duration of most manoeuvres carried out in this study did not exceed 100 s, and therefore the results obtained by means of the model that is presented may be considered to be in conformity with reality for the analysis presented in this paper.

3. Marine traffic modelling in the Gulf of Finland

3.1. Marine traffic profile

The vessel traffic profiles over the area under analysis are described using data derived from the AIS transmissions recorded in March and July 2006. The registered data do not fully reflect the existing traffic, mostly because of the limitations of the requirements for carrying AIS [43]. Regulation 19 of SOLAS Chapter V requires AIS to be fitted aboard: all ships of 300 tonnes gross and upwards engaged in international voyages, all ships of 500 tonnes gross and upwards not engaged on international voyages, and all passenger ships irrespective of size. Because of this, all “small traffic” is not included in the present study, although a large number of pleasure craft and fishing boats navigate in the Gulf of Finland, especially during the summer. These boats, despite their small dimensions, may, in some situations, complicate the traffic and raise the already high risk of collision and grounding in the area. For further research this kind of traffic should be estimated as well. Another reason behind the difference between registered traffic and the actual situation is the incompleteness of the AIS information transmitted by vessels. Transmissions without a Maritime Mobile Service Identity number (MMSI), latitude, or longitude were not stored in the database. Thus, the total number of vessels recorded in an area may be smaller than the actual number by some per cent. The area in question is the junction of two main waterways: one leads N–S and the other E–W. The N–S stream consists mainly of passenger vessels cruising between Helsinki and Tallinn, whereas the E–W stream consists of cargo vessels bound for and from harbours located in the Gulf of Finland. To estimate the number of vessels that arrive in and depart from the Gulf of Finland, counting gate number 1 was established along meridian 023.5°E. To compute the traffic volumes of the streams in the junction, another two counting gates were established. Gate number 2 was established along parallel 60°N to count N–S traffic, and gate number 3 along the meridian 026°E to count E–W traffic (Fig. 11).

The types of vessels and their percentage share of the traffic in the Gulf of Finland are presented graphically in Fig. 12. The diagram constitutes the results of 2 months of AIS transmission recordings, carried out in March and July 2006 at counting gate numbers 1 (all vessels) and 2 (passenger vessels only). The traffic recorded in March is considered the winter profile of traffic, whereas the traffic registered in July is a summer profile of marine traffic in the Gulf of Finland.

In the diagram shown in Fig. 12 the group labelled “other” consists of tugboats, icebreakers, research vessels, support vessels, sailing yachts, and vessels for which AIS transmission was not complete. In the case of messages with an MMSI number, but without vessel details, the missing information was extracted from the external ship database. This external database was very helpful in categorising vessels other than passenger ship and
tankers, whose AIS status was "cargo vessels" (container carriers, general cargo, ro-ro, bulk carriers). Afterwards the mode and maximum values of length, breadth, and draught in each class of vessel were calculated, and the results are presented in Table 5. The mode value denotes the frequently recorded values for a given class of vessel, so they may also be called typical parameters of vessels operating in the Gulf of Finland in the year 2006.

The total number of vessels registered at gates 1 and 2 is presented in Table 6. There are no significant differences in the number of vessels operating in the Gulf of Finland in the year 2006. Although the number of vessels passing gate 1 is similar throughout the year, the spatial distribution of traffic is different, which depends mostly on icing conditions in the area. As an example, histograms of the winter and summer traffic are presented in Fig. 13. The differences in the distributions are significant and cannot be omitted in risk analysis. The above relations hold true for the other gates as well.

For the purposes of safety assessment and traffic modelling the traffic spatial distributions may be approximated either by statistical distributions, or by histograms if the registered traffic fits any known distribution poorly.

### Table 5
Types and sizes of typical and maximal vessels operating in the Gulf of Finland in the year 2006.

<table>
<thead>
<tr>
<th>Vessel type</th>
<th>L_mode</th>
<th>L_max</th>
<th>B_mode</th>
<th>B_max</th>
<th>T_mode</th>
<th>T_max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ro-ro</td>
<td>162.0</td>
<td>195.0</td>
<td>20.6</td>
<td>28.1</td>
<td>6.7</td>
<td>10.0</td>
</tr>
<tr>
<td>Tankers</td>
<td>239.0</td>
<td>320.0</td>
<td>27.3</td>
<td>58.0</td>
<td>11.2</td>
<td>22.0</td>
</tr>
<tr>
<td>Container carriers</td>
<td>124.5</td>
<td>192.7</td>
<td>22.5</td>
<td>25.4</td>
<td>8.7</td>
<td>10.5</td>
</tr>
<tr>
<td>Passenger ships</td>
<td>171.3</td>
<td>266.2</td>
<td>28.7</td>
<td>36.0</td>
<td>5.0</td>
<td>8.5</td>
</tr>
<tr>
<td>General cargo ships</td>
<td>85.0</td>
<td>173.5</td>
<td>12.5</td>
<td>27.6</td>
<td>5.3</td>
<td>10.9</td>
</tr>
</tbody>
</table>

### Table 6
Number of vessels passing gate 1 (E–W traffic) and gate 2 (N–S traffic), registered in March (***) and July (*) 2006.

<table>
<thead>
<tr>
<th>Vessel type</th>
<th>Gate number 1</th>
<th>Gate number 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>East</td>
<td>West</td>
</tr>
<tr>
<td>Ro-ro</td>
<td>397</td>
<td>400</td>
</tr>
<tr>
<td>Tankers</td>
<td>618</td>
<td>607</td>
</tr>
<tr>
<td>Container carriers</td>
<td>470</td>
<td>464</td>
</tr>
<tr>
<td>Passenger ship</td>
<td>505</td>
<td>502</td>
</tr>
<tr>
<td>General cargo ships</td>
<td>895</td>
<td>929</td>
</tr>
<tr>
<td>Bulk carriers</td>
<td>153</td>
<td>145</td>
</tr>
<tr>
<td>Total</td>
<td>3038</td>
<td>3047</td>
</tr>
</tbody>
</table>

Fig. 13. Traffic flow at gate number 1 in March (to left) and in July (to right) in 2006.
the distribution. In the case of the North flow, the distribution fits well, but for the South flow, there is a significant difference. One may notice two peaks in the histograms, which are not considered in the distributions. Therefore this variable cannot be modelled by the distribution, and a random sample from recorded data is used as the input value.
3.2.2. Vessels’ length and breadth modelling

Ship lengths and breadths are modelled for the E–W and N–S streams separately; the percentage shares of the types of vessels and the distributions used are shown in Tables 7–10.

3.2.3. Modelling of lateral distribution of vessel across waterways

The distribution of vessel positions across the waterways differs and depends on season and geographical location (Fig. 13). In this research summer traffic in the junction between Helsinki and Tallinn is considered. The E–W stream is modelled on the basis of the data obtained from gate number 3 (Fig. 11), whereas the N–S stream is modelled on the basis of data from gate number 2. Data from gate number 2 were filtered and sorted to remove vessels that did not enter the crossing being analysed (ferries from Helsinki to Stockholm, and vessels not bound for Tallinn). These two streams were first described by continuous distributions and the results (number of collision candidates) were compared with the results obtained from the model, which used the distribution of vessels across the waterways in the form of discrete values. The discrepancies were significant, and it was decided not to model this variable by means of continuous distribution but use discrete values as the input and random sampling as the method of data acquisition. Histograms of considered values with the best fit distributions (to present the discrepancies) are shown in Figs. 17 and 18.

Table 7

<table>
<thead>
<tr>
<th>Vessel type</th>
<th>Percentage share in E–W flow</th>
<th>Mode (m)</th>
<th>Maximum (m)</th>
<th>Minimum (m)</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ro-ro</td>
<td>15</td>
<td>162.0</td>
<td>195.0</td>
<td>77.0</td>
<td>Triangle</td>
</tr>
<tr>
<td>Container carriers</td>
<td>17</td>
<td>124.4</td>
<td>203.0</td>
<td>47.2</td>
<td>Triangle</td>
</tr>
<tr>
<td>Passenger ships E–W</td>
<td>18</td>
<td>180.7</td>
<td>266.2</td>
<td>79.0</td>
<td>Triangle</td>
</tr>
<tr>
<td>Tankers</td>
<td>20</td>
<td>239.0</td>
<td>320.0</td>
<td>58.3</td>
<td>Triangle</td>
</tr>
<tr>
<td>General cargo ships</td>
<td>30</td>
<td>85.0</td>
<td>173.5</td>
<td>42.5</td>
<td>Triangle</td>
</tr>
</tbody>
</table>

Fig. 15. Histograms of vessels’ courses at gate number 1 with distributions fitted (eastbound traffic to the left, westbound traffic to the right).

Fig. 16. Histograms of vessels’ courses at gate number 2 with distributions fitted (northbound traffic to the left, southbound traffic to the right).

Logistic(367,2069; 5,4454)

BetaGeneral(22,273; 5,9173; 83,701; 218,35)
The distance between ships’ recorded positions and the fixed reference point for each gate was computed as follows:

- at gate 2 the position of the reference point was: Latitude 60°00.0’N, Longitude 024°51.5’E; the distance was computed along the parallel from this position towards the East,
- at gate 3 the position of the reference point was: Latitude 59°49.5’N, Longitude 026°21.6’E; the distance was computed along the parallel from this position towards the North.

### 3.2.4. Model output

The outputs of the model are: the number of encounters (crossing, overtaking, head-on), the geometrical probability of a collision for a specified type of encounter, and the mean time between accidents (collisions). Because of the complexity of the traffic over the junction, it was divided into four crossing areas, as shown in Fig. 19, and each area was analysed separately. The following main interactions between flows were analysed: crossing situations in four crossing areas, named NE, NW, SW, and SE; overtaking situations in four traffic lanes – North, South, East, and West; and head-on situations between two pairs of lanes: East–West and North–South. The length of the E–W leg was 80 Nm, and 30 Nm in the case of the N–S leg.

### 3.3. Marine accidents

A thorough analysis of accidents for the whole Gulf of Finland was conducted by Kujala et al. [6]; in their study a detailed accident statistics for the period 1997–2006 were presented and the results of theoretical models were compared with these statistics. The present analysis is limited to the area of the junction between Helsinki and Tallinn, and therefore only...
accidents that happened there are considered. Data about marine accidents in the Gulf of Finland were extracted from the HELCOM accidents database, which covers the period between 1987 and 2007. According to this database, in the junction between Helsinki and Tallinn there have been 3 collisions during the last 20 years. The positions of these collisions are listed in Table 11 and a graphical presentation is provided in Fig. 20.

It may be noticed that two collisions happened during the winter season (in the years 1996 and 2003); one of them involved an icebreaker, but there is no additional information. According to the archive reports concerning the ice cover in the Gulf of Finland issued by the Finnish Meteorological Institute, at the beginning of March 2003 the maximum ice extent was reached, and the whole Gulf of Finland was covered with ice; in 2001 the ice cover started to develop in the Gulf of Finland in late December; in February 1996 the ice cover extended from the east as far as the longitude of Helsinki, and the rest of the Gulf was ice-free. Therefore it may be assumed that the collision that takes place in the presence of ice shall be considered a winter traffic collision; otherwise it is a summer traffic collision. Thus the collision that involved the icebreaker in 2003 was a winter traffic collision, whereas the other two due to lack of ice were considered summer traffic collisions.

The positions of all collisions that happened in the period under analysis are marked in Fig. 20. The collisions that happened in the area of the junction between Helsinki and Tallinn are labelled with exclamation marks; other collisions are labelled with flags.

Drawing inferences about navigational safety from the accident statistics only or carrying out validation of models that assess the probability of a collision on the basis of number of accidents only, which are very rare, is quite difficult and highly uncertain. Therefore it seems justified to analyse the safety of navigation on the basis of the numbers of both accidents and near-miss situations, which may better reflect the collision hazard [44]. In their report Berglund and Huttunen [45] analysed the meeting situations in the Gulf of Finland for summer traffic (May, July, July) in 2006, 2007, and 2008. They provided data concerning near misses for crossing, overtaking, and head-on encounters. A near-miss situation was defined there as a meeting of two vessels being on collision courses at a distance of less than 0.3 Nm. In this paper the number of near-miss situations is roughly equivalent to the number of collision candidates computed by the model presented. In the analysis an assumption, in common use about blind navigation was made, which means that vessels do not perform any manoeuvres to avoid each other. In the case of real encounters vessels are able to steer, and therefore it may be said that the near misses are controlled, and the number of near-miss situations is affected by human factors; therefore the assumption of blind navigation is not met. Therefore, the numbers of near misses in the case of blind navigation may be expected to be even higher than those presented in the report. Abridged results of the analysis are collected in Table 12.

### 4. Results

Using the model described above, the numbers of encounters per month for three types of vessel meeting situations are obtained. Calculations were performed on the basis of the summer traffic profile, based on AIS data recorded in July 2006. The modelled values are compared with the number of encoun-
ters observed [45], number of encounters computed by means of Pedersen’s model [8], and number of encounters obtained with modified Pedersen model in which the collision diameter values were replaced with MDTC values. The results of this experiment for all four crossing areas according to Fig. 19 are shown in Table 13.

In the next step the probability of a collision is computed on the basis of the number of encounters obtained and the causation probability values adopted. This paper uses two causation probability values are commonly accepted [6–8]: 1.3 × 10⁻⁴ for crossing courses and 4.9 × 10⁻⁵ for parallel courses. The obtained values of the probability of a collision and the time between collisions for each crossing area are presented in Table 14.

The differences between the estimated and observed values are presented in Fig. 21. They are expressed as a percentage of observed values according to the formula

\[ D = \left( \frac{N_{\text{EST}} - N_{\text{OBS}}}{N_{\text{OBS}}} \right) \times 100 \% \]  

(27)

where \( N_{\text{EST}} \) denotes the number of encounters estimated by means of the appropriate model and \( N_{\text{OBS}} \) is the number of encounters observed.

For all the meeting cases analysed the presented model overestimates the number of encounters in comparison with the observed values in the case of crossing situations by 19%, whereas Pedersen’s model underestimates this number by almost 70%. The number of encounters computed by a modified Pedersen model, with an MDTC module applied, is still 27% less than the observed value.

In the case of overtaking the presented model overestimates the result by 27%, while, on the contrary, the results given by Pedersen’s model are underestimated by 54%.

The numbers of head-on meetings determined by the MDTC based model and the Pedersen’s model are similar, and in both cases the results are overestimated, by 140% and 128%, respectively. In both cases only E–W traffic was considered, and N–S traffic was left out. The assumption was made that vessels cruising in the N–S flow, which are passenger liners, are operated by highly skilled crews, and head-on meetings do not occur (and are controlled if they do).

In the case of overtaking, the presented model overestimates the result by 27%, while, on the contrary, the results given by Pedersen’s model are underestimated by 54%.

The numbers of head-on meetings determined by the MDTC based model and the Pedersen’s model are similar, and in both cases the results are overestimated, by 140% and 128%, respectively. In both cases only E–W traffic was considered, and N–S traffic was left out. The assumption was made that vessels cruising in the N–S flow, which are passenger liners, are operated by highly skilled crews, and head-on meetings do not occur (and are controlled if they do).

The overestimation of the results given by the MDTC based model seems natural. In real traffic, ships have the opportunity to avoid each other; therefore the assumption of blind navigation, which is the principle of the geometrical model, is not met. Because of this discrepancy one may expect models to give larger results than those that can be gained by the observation of real traffic.

### Table 12

<table>
<thead>
<tr>
<th>Encounter/Number</th>
<th>Head-on</th>
<th>Crossing</th>
<th>Overtaking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>8.0</td>
<td>419.0</td>
<td>943.0</td>
</tr>
<tr>
<td>Per month</td>
<td>2.8</td>
<td>142.7</td>
<td>321.1</td>
</tr>
</tbody>
</table>

### Table 13

<table>
<thead>
<tr>
<th>Data source Type of meeting</th>
<th>Crossing</th>
<th>Overtaking</th>
<th>Head-on</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berglund and Huttunen [45]</td>
<td>142.7</td>
<td>321.1</td>
<td>2.8</td>
</tr>
<tr>
<td>MDTC based model</td>
<td>169.0</td>
<td>407.0</td>
<td>6.8</td>
</tr>
<tr>
<td>Pedersen</td>
<td>46.8</td>
<td>146.4</td>
<td>6.4</td>
</tr>
<tr>
<td>Pedersen_MDTC</td>
<td>104.3</td>
<td>146.4</td>
<td>6.4</td>
</tr>
</tbody>
</table>

### Table 14

<table>
<thead>
<tr>
<th>Type of encounter</th>
<th>Probability of collision per year</th>
<th>Number of candidates per year</th>
<th>Time between collisions in years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crossing</td>
<td>NE 0.057</td>
<td>440</td>
<td>17.6</td>
</tr>
<tr>
<td></td>
<td>NW 0.077</td>
<td>593</td>
<td>13.0</td>
</tr>
<tr>
<td></td>
<td>SW 0.054</td>
<td>415</td>
<td>18.5</td>
</tr>
<tr>
<td></td>
<td>SE 0.075</td>
<td>576</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.263</td>
<td>2024</td>
<td>3.8</td>
</tr>
<tr>
<td>Overtaking</td>
<td>Traffic lane E 0.120</td>
<td>2568</td>
<td>8.0</td>
</tr>
<tr>
<td></td>
<td>W 0.144</td>
<td>2304</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>N 0.009</td>
<td>180</td>
<td>113.0</td>
</tr>
<tr>
<td></td>
<td>S 0.009</td>
<td>180</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.282</td>
<td>5232</td>
<td>3.5</td>
</tr>
<tr>
<td>Head-on</td>
<td>Traffic lane E–W 0.004</td>
<td>82</td>
<td>249.0</td>
</tr>
<tr>
<td></td>
<td>N–S 0.033</td>
<td>667</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.037</td>
<td>749</td>
<td></td>
</tr>
<tr>
<td>Grand total</td>
<td>0.582</td>
<td>8005</td>
<td>1.7</td>
</tr>
</tbody>
</table>

---

**Fig. 20.** Positions of ship–ship collisions in the Gulf of Finland between years 1987 and 2007.
5. Conclusions

This paper presents a new approach for collision probability modelling, which is an important part of risk analysis process. The innovative use of a ship motion model to determine the Minimum Distance To Collision may be recognised; this approach meets the demands expressed in recent works and the obtained results are very promising. The number of encounters in the waterways separated by Traffic Separation Schemes; these factors, among others, may significantly reduce the causation probability. At present research is being carried out at Aalto University that aims to redefine the factors that affect the causation probability. They will help to tailor the appropriate causation probability model for the model proposed.

The model presented here is based on MDTC values, computed with the assumption that two vessels take action designed to avoid collision. The analysis of MDTC in the case of single-ship manoeuvres is the aim of the next step of this research.

Another essential factor that should be taken into account in further research is the variation in marine traffic according to the time of the day. In the model presented here intensity of ships is assumed to be constant during the day, and peak hours are not taken into account, whereas they exist both in the N–S and the E–W streams. Although this factor would not change the total number of collision candidates, it might help define the high-risk periods during the day.

The model presented should also be verified in other sea areas for which near-miss data are available. As more AIS data become available, successive analyses for the years 2008 and 2009 are expected to be carried out.

Acknowledgements

D.Sc. Erik Sonne Rav from the Danish Maritime Safety Administration in Copenhagen is thanked for his help in evaluating the traffic flows in the Gulf of Finland on the basis of the AIS data.

D.Sc. Jarosław Artyszuk from the Maritime University of Szczecin in Poland is thanked for his invaluable comments on ship motion model adaptation.

The authors also appreciate the financial contributions of the following entities: the EU, Baltic Sea Region (this research was carried out within EfficienSea project), the city of Kotka, and the Finnish Ministry of Employment and the Economy.

References


