



RESEARCH ARTICLE

Method for prediction of ship traffic behaviour and encounter frequency

Hiroko Itoh 

National Maritime Research Institute, National Institute of Maritime, Port and Aviation Technology, Tokyo, Japan.
Corresponding author. E-mail: hiroko@m.mpat.go.jp

Received: 22 December 2020; **Accepted:** 08 September 2021; **First published online:** 19 November 2021

Keywords: future traffic prediction; encounter frequency; traffic rules

Abstract

The design of new rules on seaways, such as traffic restrictions, requires determining the degree of improvement in marine traffic safety beforehand by considering the occurrence of new hazardous factors. This study proposes a method to predict the future traffic behaviour and ship encounter frequency (EF) with the introduction of a new traffic rule. First, a sensitivity analysis is conducted to identify the factors affecting the EF. A method of predicting future traffic behaviour and EF is presented based on the analysis of changes in the traffic flow in an area with a temporal restriction. Results show that the method appropriately predicts the location and degree of the occurrence of encounters in the sea area. The proposed method contributes to the discussion of future traffic safety, when sailing in a specific area is restricted by new regulations, installations of new offshore wind farms and fishing reefs.

1. Introduction

Ship accidents in coastal areas generally have a more adverse impact than do those in ports and inland waters (Japan Coast Guard (JCG), 2018). Collision accidents, the predominant type of accidents involving fatalities, account for approximately one-fourth of ship accidents (European Maritime Safety Agency (EMSA), 2019). One safety measure to prevent collisions is to decrease the number of encounters via maritime traffic management, which separates ship traffic based on the direction of travel. Since the 1970s, when the International Maritime Organization (IMO) introduced this measure, routeing systems, such as traffic separation schemes (TSSs), have been established in coastal waters around the world (IMO, 2003).

Although there are many congested sea areas around Japan, there are only a few voluntarily separated navigation areas, except for narrow waters. A reason for the lack of routeing systems in the coastal waters of Japan is that the sea is widely recognised as a common property of various industries, such as fishing and leisure. Therefore, to implement a new traffic management measure, the parties concerned must reach a consensus, which remains an impediment.

In 2016, the JCG deemed the preservation of safety in coastal areas with heavy traffic as an immediate concern and proposed to the IMO a recommended route off the western coast of Izu O Shima Island, which is located at the entrance of Tokyo Bay. The proposal document includes the results of the preliminary analysis, the developed management measures and the selection of the most suitable measure for this area (IMO, 2016). The author was part of the team commissioned to design the recommended route when the proposal was being prepared. Preliminary analysis of the present traffic conditions and historical accident data revealed that most collisions occurred due to head-on conflicts owing to mixing of the traffic to and from Tokyo Bay. Accordingly, the team drafted recommended routes to reduce such

collisions, predicted the future traffic under each option and estimated their safety and economic effects (Miyake et al., 2016, 2017).

This study proposes a systematic method called EnFreq that was conceived to estimate and predict the ship encounter frequency (EF) in such drafting tasks. The remainder of this paper is organised as follows. Section 2 summarises the basic concepts of the EF and describes various analysis techniques relating to the EF. Section 3 describes the characteristics of the target area, which is used to validate the proposed method. Section 4 presents the sensitivity analysis conducted to identify the variables that significantly influence the EF estimation. Subsequently, this section demonstrates the method of predicting the EF based on these identified variables, along with the predicted EF distributions. Section 5 presents the conclusions.

2. Basic concept and methods

To implement effective traffic management, it is necessary to determine the location and frequency of encounters, develop safety measures, and predict the future traffic behaviour under these measures. The effectiveness of the developed safety measures can be determined by estimating the location and frequency of the encounters based on the prediction.

2.1. Estimation method for number of encounters

The methods used to identify dangerous encounters can be roughly classified into two types: deterministic and probabilistic. Deterministic methods differentiate possible collisions from various encounters involving multiple objects. Some methods to detect the collision risk states from individual combinations have been proposed for the automotive and aviation fields (Lefèvre et al., 2014; Zou et al., 2021). Over several decades, deterministic methods have been developed for providing real-time support to maritime ship operators. Therefore, their primary application involves dynamically detecting the ships moving on a course approaching a specific ship (own-ship) and estimating the degree of situational difficulty.

Among these methods, the ship domain concept has been used to describe an inaccessible area around a ship in studies of traffic capacity (Fujii and Tanaka, 1971). The obstacle zone by target (OZT) method is another deterministic method that helps in detecting the future positions of other ships which may block the own-ship's course (Imazu, 2017). In addition to providing navigational support, it is used for analysing hazardous locations. For example, in previous studies, OZT was calculated for all ship combinations in a target area to detect hazardous locations and their degrees (Miyake et al., 2017; Itoh and Miyake, 2019). Studies have reported that the calculated metrics and actual frequency are not sufficiently consistent when considering ship-to-ship collisions in sea areas (Goerlandt and Kujala, 2014).

Probabilistic methods obtain the frequency of collisions calculated from the number, position, and velocity of a group of objects. Probabilistic collision risk modelling has been proposed for aviation, space, and pedestrian and bicycle traffic management (Blom et al., 2003; Wang et al., 2018; Netjasov, 2020; Chan, 2021). Probabilistic methods for ship traffic are based on the multiplication of the number of dangerous encounters of the ships with the failure probability of the evasive manoeuvres during the encounters (Fujii, 1983). These methods are still widely used with various improvements (e.g., Pedersen, 1995; Nyman, 2009; Khaled and Kawamura, 2015; Cucinotta et al., 2017; Kawashima and Itoh, 2019). In this concept, the number of collisions, N_c , is formulated as follows:

$$N_c = P \cdot N_G, \quad (2.1)$$

where P is the causation probability and N_G is the number of possible collisions.

P represents the probability that the evasive manoeuvre fails due to human error, equipment failure and other causes. Some studies have obtained the P value by calculating the ratio of the actual number of accidents obtained from historical data and the number of possible accidents obtained from traffic

data (Fujii et al., 1984; Kawashima and Itoh, 2019). Another approach involves methods such as fault trees and Bayesian networks to derive the failure probabilities by integrating the necessary elements associated with the operator's cognitive tasks in evasive manoeuvres (Itoh et al., 2007; Martins and Maturana, 2010; Pedersen, 2010).

The number of dangerous encounters is represented by the number of possible accidents, N_G . Essentially, it represents the number of encounters wherein ships would collide if the evasive manoeuvres were unsuccessful. When two groups of ships, i and j , sail during time, T , in a certain region, S , N_G is represented as follows:

$$N_G = \rho_i \rho_j V_r D_{ij} S T, \quad (2.2)$$

where ρ_i and ρ_j are the traffic densities of groups i and j , respectively, V_r is the relative velocity and D_{ij} is the cross section of the two groups. The density, ρ , of each ship group is defined as $\rho = Q/(VW)$, where Q is the traffic volume, V is the average speed of the ships and W is the width of the waterway. The cross section, D_{ij} , is the total length of the projections of the two ships when a ship of group i is placed at the centre and a ship of group j slides completely around it. For example, if the ship width is $1/6$ the ship length, D_{ij} takes a value between $(L_i + L_j)/6$ and $(L_i + L_j)$, depending on the crossing angle of the two groups (Matui et al., 1983). In the coastal waters of Japan, most collisions are caused by ships travelling in opposite directions (Itoh et al., 2012). When the angle between the paths is 180° , D_{ij} reaches its minimum value.

The concept of the number of potential accidents is often used in collision-frequency estimation studies. It is also known as the number of collision candidates, which represents the number of times two ships have entered a collision course (Montewka et al., 2010; Pedersen, 2010; Silveira et al., 2013). Apart from directly using physical quantities such as position and velocity, their probability distributions can also be used as inputs. The resulting value can be considered as the probability distributions of the number of potential accidents.

From the perspective of risk assessment, it has been observed that these methods do not adequately consider the navigation method of the actual ship; thus, they lack justification and present uncertain results (Goerlandt and Kujala, 2014). Additionally, the calculation only includes the conflicts between two ships and does not consider simultaneous encounters involving three or more ships. Nevertheless, this approach is adopted because it meets the requirements of this study, i.e. it helps in determining the location and frequency of encounters occurring in the sea area, with most of the conflicts occurring between two ships.

2.2. Traffic model for estimation of number of encounters

When the number of encounter estimation formula, i.e. Equation (2.2), is used, the information on the ships sailing in the target area is required as an input. The traffic data obtained by radar or automatic identification system (AIS) can be used directly or can be reproduced using a probability distribution function.

To provide these inputs, some traffic models comprise ships and fairways, which, for example, include geometrical representations of fairways formed by the shapes of connected lines called legs or route lines (Christensen et al., 2001; Friis-Hansen, 2008; DNV-GL, 2015). When many ships have similar tracks, the data can be handled easily by defining a fairway and associating them with it. For example, IWRAP provides a function to estimate the collision frequency in a waterway by defining a leg for each of the obtained track groups (Friis-Hansen, 2008).

Conversely, when there are multiple origins and destinations near each other in the area or when the ship tracks differ based on the ship properties, route identification becomes complicated. To address this issue, Kawashima et al. (2015) proposed a method to automatically detect a group of ships with similar track lines in numerous ship tracks using principal component analysis. Seshita et al. (2016) developed a method to automatically generate a traffic flow tube model which represents a group of similar tracks

using a clustering algorithm. In this study, ships with the same origin and destination locations were considered as belonging to the same group.

2.3. Estimation method of EF

Recently, maritime traffic surveys have been conducted based on AIS ship track data owing to the prevalence of the AIS systems onboard (Montewka et al., 2010; Tsou, 2010; Silveira et al., 2013; Weng and Xue, 2015; Altan and Otay, 2017; Mujeeb-Ahmed et al., 2018). This is because the AIS simplifies the method of observing a broader region for an extended period. When a large amount of track data is available, the number of dangerous encounters can be calculated by dividing it into small areas and short time periods. Specifically, the number of encounters of Equation (2.2) per unit area and time is defined as EF (Kawashima and Itoh, 2019). In this study, EF between groups i and j per unit area and time, $E_f(i, j)$, is defined as follows:

$$E_f(i, j) = \rho_i \rho_j V_{ij} D_{ij}. \quad (2.3)$$

The EF concept provides a simplified perspective of the encounters; even in complex traffic with many fairways, tracks can be divided into simple traffic in a small region. Based on this, Kawashima and Itoh (2019) calculated N_G for all the combinations of the course angles for traffic near the entrance of Tokyo Bay to identify the locations with dangerous encounters. For the sake of simplicity, this study focuses on the encounters that occur between the southbound and northbound traffic in the opposite directions. The data required for Equation (2.3) was obtained from the values in each unit area.

2.4. Traffic behaviour prediction

Some studies have predicted ship behaviour under new restrictions. Christensen et al. (2001) represented the distribution of the ship traffic using a Gaussian distribution and predicted the changes after the installation of an offshore wind farm. Szlapczynski (2013) proposed a search method to determine an optimal safe track in an area with a navigation separation zone. Pietrzykowski et al. (2015) reported that most ships navigate according to the regulations concerning the TSSs, which makes their behaviour predictable.

The factors of the traffic at each point in the area vary with the behaviour of each ship according to the rules. The factor changes to be predicted depend on the application purpose, i.e. the estimation function in Equation (2.2) in this case. Therefore, identifying the influential variables in the function is essential for practical and accurate prediction. Sensitivity analysis detects the magnitude of influence of each variable. Ylitalo (2009) conducted sensitivity analysis of a collision model for traffic in the Gulf of Finland and observed that the causation probability and traffic volume were the most sensitive variables.

Most of these studies were conducted for individual aims of predicting traffic behaviour or determining influential variables. They did not analyse the traffic safety around the restricted areas and did not provide a systematic method to predict the collision frequency. This study proposes a systematic method to predict the changes in traffic behaviour due to restrictions. It is based on a preliminary analysis of the traffic conditions along with a sensitivity analysis. It does not consider any transient response, including the collision avoidance manoeuvres and the avoidance of stormy weather.

3. Traffic in the survey area

To understand the ship behaviour associated with restrictions in fairways, an area with relatively simple traffic is considered off the east coast of Fukushima Prefecture in north-eastern Japan. This area is selected because of the rules that were in place to restrict navigation in the past.

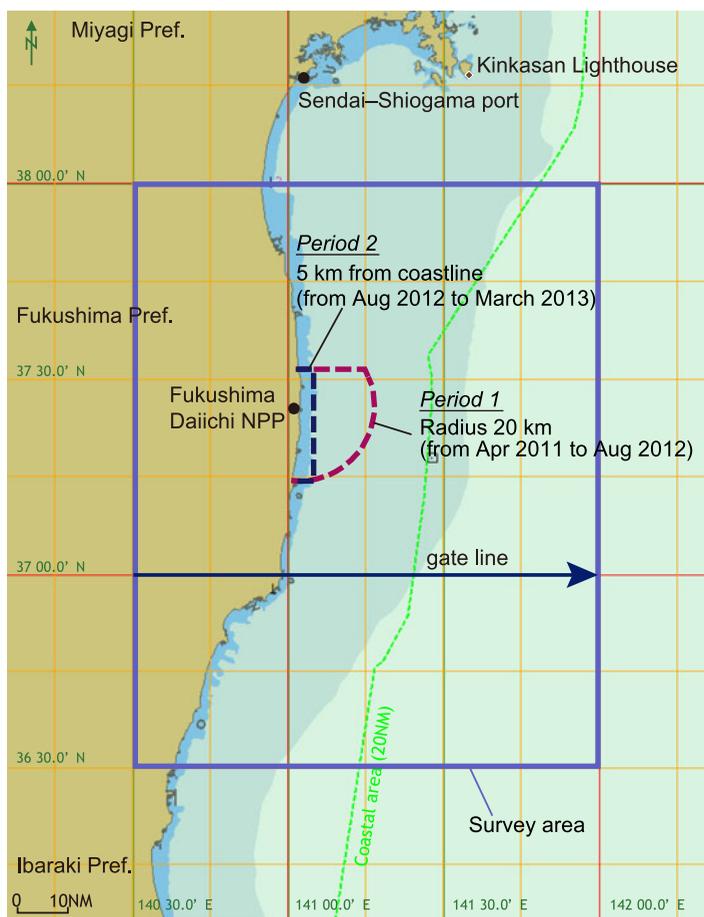


Figure 1. Survey area.

3.1. Survey area

Figure 1 shows a map of the survey area. The primary traffic in this region runs almost parallel to the land, which is common in coastal waters. After the Great East Japan Earthquake on 11 March 2011, access to the surrounding area within 20 km was limited between April 2011 and August 2012 due to the accident that occurred at the Fukushima Daiichi Nuclear Power Plant. According to Fukushima Prefecture and the JCG, the evacuation order zone has been settled and resized several times after the disaster (JCG, 2012; Fukushima, 2019).

This paper focuses on two of these periods: one is when the evacuation order zone occupied 20 km from the coast, and the other when it was sufficiently small. Table 1 presents an overview of these periods. In this area, the water depth is sufficient at a distance of more than 5 km from the coast. Conversely, the restricted sea area in Period 2 includes shallow water and is rarely used by merchant ships. Therefore, Period 2 is considered virtually nonrestrictive, and Period 1 is considered a duration with a restriction. The ship track data for one month are utilised. Note that the data used in some of the analyses do not include the first and last days because they require the departure/arrival locations, which are identified by the tracks of the previous/next day. Otherwise, all the data are used. In this study, ships equipped with class A AIS are targeted. However, in principle, the same analysis can be performed for ships equipped with class B AIS and those observed by radar, if the information required for the corresponding analysis is available.

Table 1. Summary of the periods and evacuation order zones.

	Period 1	Period 2
Evacuation order zone	Bounded by the latitude at 37° 30' 49.6'' N, the shoreline, and a 20-km-radius-arc centred on the plant	Bounded by the latitudes at 37° 30' 49.6'' N and 37° 18' 59'' N, the longitude at 141° 5' 20'' E, and the shoreline
Applicable period	16 April 2012–9 Aug. 2012	10 Aug. 2012–5 March 2013
Period of applied data	1–31 July 2012 (31 days)	1–31 Dec. 2012 (31 days)
Period of data used in analysis	2–30 July 2012 (29 days)	2–30 Dec. 2012 (29 days)

Table 2. Traffic volume (number of ships) by ship type (29 days).

	Period 1		Period 2	
Number of ships	2320	100%	2174	100%
<i>Merchant ships</i>	2215	95%	2062	95%
Cargo	1343	58%	1191	55%
Tanker	739	32%	745	34%
Passenger	133	6%	126	6%
<i>Other ships</i>	105	5%	112	5%
Tugboat	21	1%	34	2%
Fishing	7	0%	5	0%
Other	77	3%	73	3%

3.2. Traffic overview

To provide an overview of the traffic conditions during the periods, the ship track data are obtained when travelling through the east–west gate line at 37.0°N latitude, arranged in full width. Table 2 shows the traffic volume based on the ship type. Ship types are categorised by the ‘Type of Ship’ number in the AIS data (ITU, 2014). Specifically, 70, 80 and 60 indicate cargo ships, tankers and passenger ships, respectively. Additionally, 31, 32 and 52 indicate tugboats, whereas the other 30 represent fishing boats. As shown in Table 2, cargo ships account for the most significant proportion (more than half), followed by tankers and passenger ships. The total proportion of the tugboats, fishing boats and other types is less than 6%. There is no significant variation in this trend over time.

Figure 2 shows the daily traffic volume based on the direction of travel during each period. The average numbers of ships per day are 78.6 and 74.4. From this figure, it is obvious that the total number of ships in Period 2 in Table 2 is less than that in Period 1, which is mainly due to the year-end holidays (year-end and New Year holidays are essential annual holidays in Japan, during which the number of domestic ships decreases). Empirically, the number of ships per day varies based on the day of the week, season, weather and economic conditions, but the data did not show significant variation. Additionally, the data provided by the Japan Meteorological Agency confirmed that no typhoons or storms interfered with the ship operations during these periods (JMA, 2019).

3.3. Geographical paths

The ship tracks were obtained by separating the AIS position data according to the ship and arranging them in a time-series sequence. The tracks were then grouped based on the origin–destination (OD)

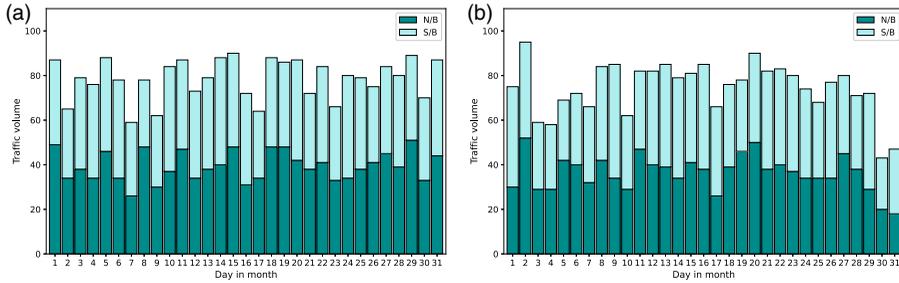


Figure 2. Traffic volume by duration and direction (31 days). (a) period 1, (b) period 2.

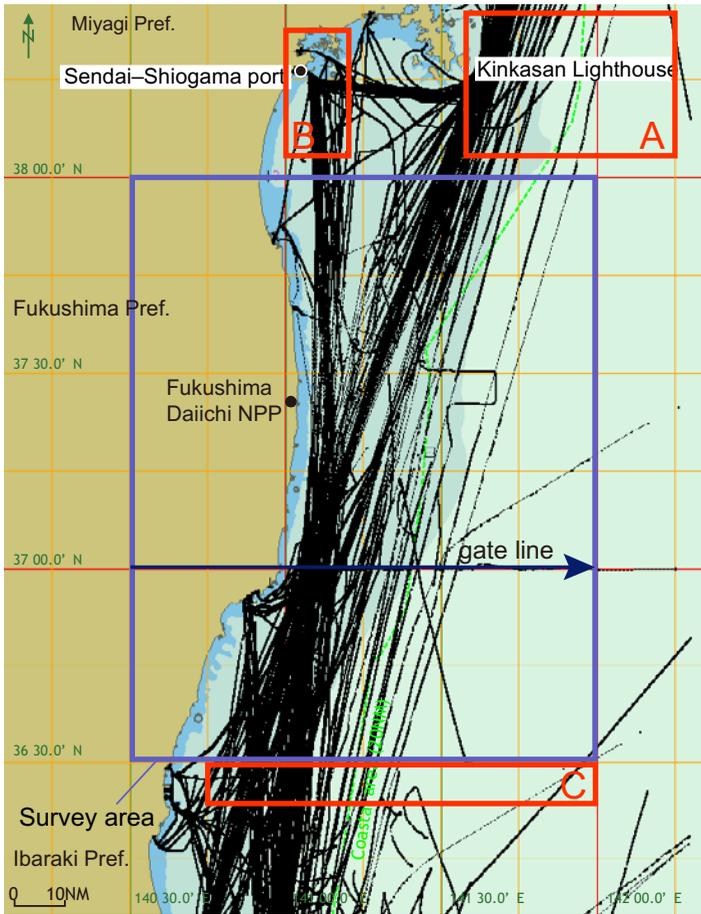


Figure 3. Ship tracks and destination areas (31 July 2018).

pairs. Figure 3 shows the recent tracks in the vicinity of the target area. Some missing values were found in the data for a particular part of the area. However, linear interpolation could be performed consistently.

The main fairway in this area diverges into two branches in the north: one towards the area off Kinkasan (Mt. Kinka) (marked with ‘A’ in Figure 3) and the other to the Sendai–Shiogama port or nearby ports (marked with ‘B’). In the southern part of the area, the tracks merge to overlap at the southern end (marked with ‘C’). In summary, the northbound and southbound paths of the two main groups with common OD pairs were grouped as geographic paths to be analysed as the subjects of this study.

4. Analysis

This section describes the method of predicting the change in the traffic behaviour according to the restrictions. As a preliminary analysis, the effects of traffic factors on the number of encounters are clarified using a simple sensitivity analysis. The differences in the traffic behaviour between the two periods are then presented, and a method of predicting the traffic behaviour is described.

4.1. Sensitivity analysis

From Equation (2.3), the number of encounters, E_f , depends on three variables: Q , V and L . Furthermore, the range of values they can take depends on the water. The effects of these variables are identified by the following steps:

1. The target sea area is divided into small grids, which are created by arranging gate lines at regular intervals. Depending on the direction of traffic, the gate lines are set eastwards from 140.5°E longitude. The grid size, S , is approximately 1850 m \times 1484 m (1/60° of the longitude and latitude at this location). The total number of grids is 90 \times 90.
2. In each small grid, the data of the above three variables are acquired in two directions. Information on all the ships is collected at each gate line. The target period is Period 1 (29 days).
3. E_f for all the grids is calculated by changing the values of the three variables within the range of the acquired data. D_{ij} is determined first. As the angles formed by the mean courses over ground (COG) of the pairs of groups are approximately 180°, the cross section is set to $D_{ij} = (L_i + L_j)/2$. Equation (2.3) is then used to calculate E_f by varying the values of Q , V and L individually. Considering that the properties of the probability distributions of the density, velocity, and length are highly biased, the medians are applied as the base case. The upper and lower bounds of each distribution are applied to the 95th and 5th percentiles of the values obtained by the AIS, respectively.

Table 3 shows the settings and results of the sensitivity analysis. The calculations were performed using the International System of Units, and the results were then multiplied by S and converted into years. Variable Q represents the number of ships per second, V is the average velocity of the ships converted to m/s, and L is the average length of the ships. Subscripts i and j indicate each of the two ship groups (north- and southbound, respectively). The $E_f \cdot S$ value of 1.2e-02 [times/year] for the base case represents the number of encounters per year in a single grid, where all the variables take the median values. Setting Q to the upper bound yielded an $E_f \cdot S$ value of 1.0 [times/year]. Incidentally, the total number of encounters of the ships in the opposite directions, i.e. the total $E_f \cdot S$ of all the grids in the target area, was approximately 921 times/year.

The sensitivity analysis results indicate that the influence of variable Q is prominent, being 40 times or more compared with that of V and 30 times or more compared with that of L . This indicates that the traffic density is the most influential variable in estimating the EF. In summary, accurate estimation or prediction of high-density locations and frequency is the most important factor in identifying locations where encounters occur frequently and their degree in the target area.

4.2. Traffic change accompanying the change in controlled waterway

Traffic behaviour in coastal areas is generally rational. Many trajectories are linear, connecting waypoints not far from the coastlines (Itoh and Yakabe, 2014). If there are traffic restrictions on such routes, ships will consequently avoid the restricted areas. This study considers the behaviour of ships in such cases.

Figure 4 shows the traffic density distributions for both the periods. The shape of the high-density regions, which is essential for calculating EF, is similar to a passage connecting waypoints in a straight line. It is obvious that the high-density regions are concentrated in a narrow width of less than 10 nautical miles in the lateral direction. Section 4.3 details these observations. Additionally, it can be observed that the ships generally follow the restrictions for their respective periods and avoid the restricted areas.

Table 3. Variable setting and results of sensitivity analysis. The columns of $Dir1(i)$ and $Dir2(j)$ show the values applied to each variable. The columns of $E_f \cdot S$ shows the resulting number of encounters in the respective period.

Variable	Unit	Percentile	Dir1 (i)	Dir2 (j)	$E_f \cdot S$		Ratio to base case	
					per second	per year		
base case					3.9e-10	1.2e-02	1.000	
Q_i, Q_j	1/s	Lower-bound	5%	4.0e-07	4.0e-07	7.0e-12	2.2e-04	0.018
			25%	8.0e-07	8.0e-07	2.8e-11	8.9e-04	0.071
		median	50%	3.2e-06	2.8e-06	3.9e-10	1.2e-02	1.000
		75%	1.3e-05	1.2e-05	7.0e-09	2.2e-01	17.679	
		Upper-bound	95%	2.6e-05	2.7e-05	3.2e-08	1.0	80.143
V_i, V_j	m/s	Lower-bound	5%	4.6	5.2	5.7e-10	1.8e-02	1.455
			25%	6.3	6.5	4.3e-10	1.4e-02	1.101
		median	50%	6.9	7.2	3.9e-10	1.2e-02	1.000
		75%	7.7	8.0	3.6e-10	1.1e-02	0.901	
		Upper-bound	95%	9.1	9.2	3.1e-10	9.6e-03	0.774
L_i, L_j	m	Lower-bound	5%	76.5	78.3	2.4e-10	7.7e-03	0.617
			25%	101.7	102.1	3.2e-10	1.0e-02	0.813
		median	50%	124.1	126.6	3.9e-10	1.2e-02	1.000
		75%	156.0	159.1	5.0e-10	1.6e-02	1.257	
		Upper-bound	95%	239.5	246.0	7.6e-10	2.4e-02	1.937

However, evasion occurs in a manner that the restricted area is barely avoided, and instances where ships sail slightly inside the restricted area can be observed.

Figure 5 shows the average COG of the four paths. At the north and south ends of the area, the difference in the COG between the periods is small for all the paths, which is consistent with the fact that no significant difference exists in the density distribution in Figure 4. A difference in the COG depending on the period is observed around the restricted area in the central part of the target area, particularly in paths BC and CB, which go back and forth between Sendai–Shiogama and the southern end. For paths AC and CA off Kinkasan, the difference in the COG depending on the period is small, which is attributed to the small overlap between the route and restricted area.

4.3. Prediction of EF

Figure 6 shows a schematic of the EnFreq prediction method. First, an estimation model for unrestricted conditions, called a ‘reference model’, is developed from the observation data. A predictive model is then created based on this model. In the prediction model, the items less affected by restrictions, such as traffic volume and ship type, and those with low sensitivity to encounters, such as length and velocity, as observed in Sections 3.2, 4.1 and 4.2, are used without modifying the values in the reference model.

Conversely, the items that are affected and have high sensitivity, i.e. the distribution of traffic density within the target area, must be predicted. To obtain this distribution, it is necessary to predict the passage–position distribution in detail. Therefore, a method is developed to predict this distribution by probabilistically modelling the distribution of the transverse passage position on the path. Moreover, as

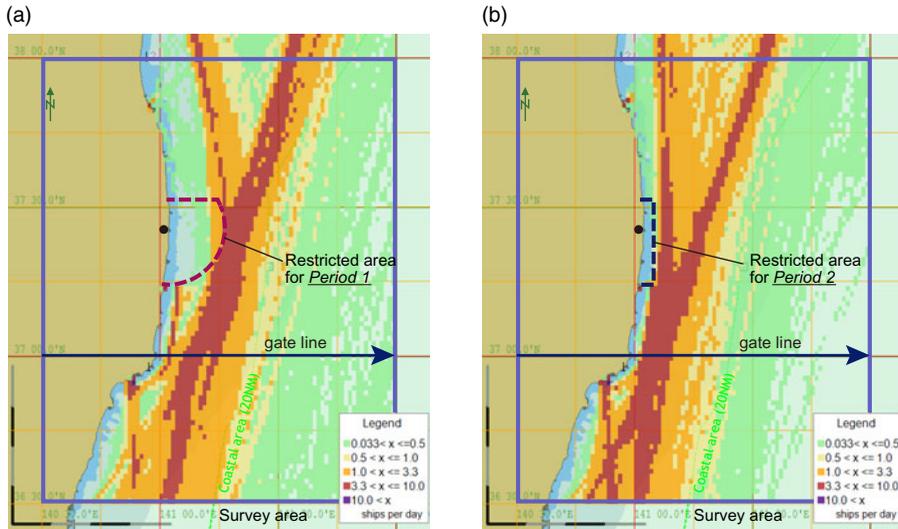


Figure 4. Traffic density before and after the change of controlled waterway. (a) period 1 (before the change), (b) period 2 (after the change).

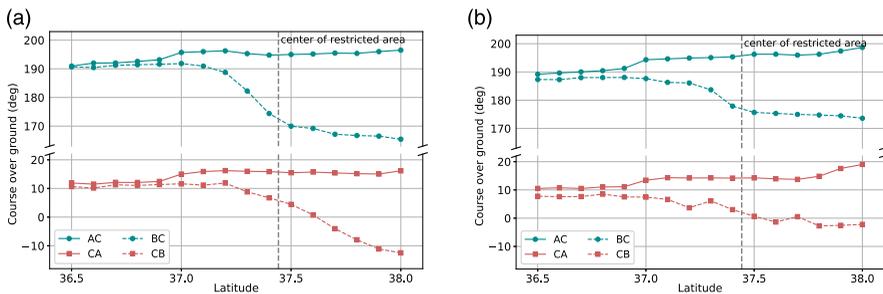


Figure 5. Mean COG of the four paths. (a) period 1, (b) period 2.

mentioned in Section 4.1, the traffic density data must ensure the accuracy of the high-density locations and their densities. Section 4.2 indicates that in the direction of travel, the high-density area is similar to a narrow line connecting the waypoints. To represent the transverse distribution of passage positions in a way that these conditions are met, characteristics of passage position distributions are identified next.

The filled bar in Figure 7 shows the observed transverse distribution of the passing positions of the ships sailing along path AC near the restricted area. The shape is asymmetric, and the kurtosis is relatively large. Multiple functions represent such a distribution. Based on a previous study (Itoh and Yakabe, 2014), a three-parameter gamma distribution function was introduced in this study (Kübler, 1979; Nagatsuka and Balakrishnan, 2012; NIST, 2012a, 2012b). The standard formula of the three-parameter gamma distribution is as follows:

$$f(x; \alpha, \beta, \gamma) = \begin{cases} \left(\frac{x-\gamma}{\beta}\right)^{\alpha-1} \frac{\exp\left\{-\frac{x-\gamma}{\beta}\right\}}{\beta\Gamma(\alpha)}, & x > \gamma \\ 0, & \text{otherwise} \end{cases} \quad (4.1)$$

where $\alpha > 0, \beta > 0, -\infty < \gamma < \infty$. Here, $\Gamma(\alpha)$ is the Gamma function, that is,

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt. \quad (4.2)$$

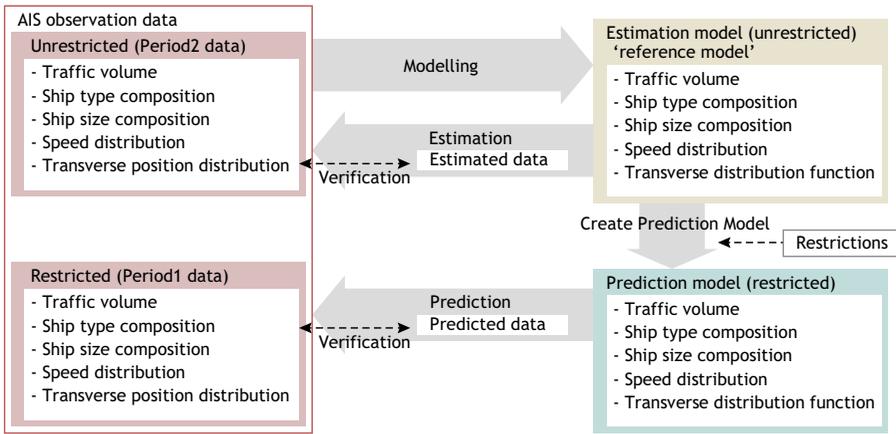


Figure 6. Schematic of the EnFreq prediction method.

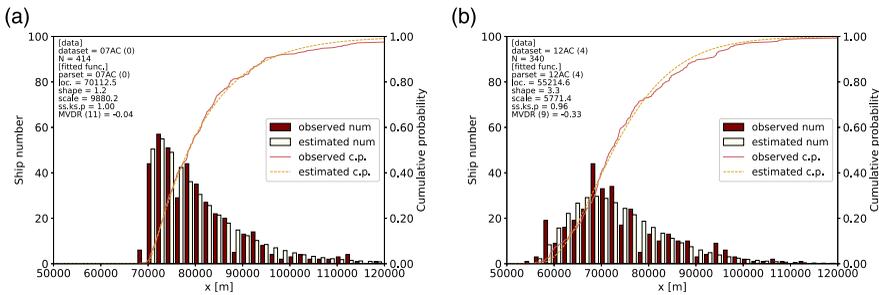


Figure 7. Observed and estimated density distributions of transverse position on path AC. (a) period 1, (b) period 2.

Cost functions, such as the least square error and least absolute error, are generally used to fit entire distributions. In this study, it is necessary to specifically fit the areas where traffic is concentrated. Accordingly, the maximum distance between the observed and measured values of the cumulative distribution is used as the cost function.

Given n observational data, x_1, x_2, \dots, x_n , sorted in ascending order, the empirical cumulative distribution function is defined as:

$$S_n(x) = \frac{1}{n} \sum_{i=1}^n X_i(x), \tag{4.3}$$

where $X_i(x) = 1(x_i \leq x)$, $0(x_i > x)$. The cost function is

$$\text{cost} = \max_{-\infty < x < \infty} |S_n(x) - F(x)|, \tag{4.4}$$

where F is the cumulative distribution of the estimated data.

To evaluate the obtained distribution function, a method for evaluating the accuracy of the obtained distribution function in representing the original distribution is introduced. This accuracy is determined in terms of whether it fits the purpose of this study and its general goodness-of-fit. From the latter perspective, the Kolmogorov–Smirnov test was introduced, which typically tests the goodness-of-fit for two large datasets. From the former perspective, an original method was developed to analyse the accuracy of the location and volume of traffic concentration. To calculate this, the distances of the passage positions from one end point, which is defined as the origin, are divided at regular intervals along the gate, which is considered as the axis. The traffic volume for each divided class is then aggregated to create a histogram to calculate the following metrics:

Table 4. Estimation result (period 1 and period 2).

Period	Path	Area	N	p-Val. (KS)	MCG	MVDR
1	AC	Central	414	1.00	0	-0.04
1	CA	Central	367	0.95	0	-0.27
1	BC	Central	220	0.89	-1	-0.06
1	CB	Central	226	0.98	-1	-0.01
2	AC	Northern end	371	0.27	0	-0.45
2	CA	Northern end	321	0.27	0	-0.40
2	BC	Northern end	230	0.91	1	-0.01
2	CB	Northern end	220	0.97	-2	0.10
2	AC	Central	340	0.96	0	-0.33
2	CA	Central	300	0.99	-2	-0.26
2	BC	Central	219	0.75	0	-0.15
2	CB	Central	208	0.93	-1	-0.09
2	AC	Southern end	364	0.63	0	-0.32
2	CA	Southern end	316	0.97	-2	-0.18
2	BC	Southern end	235	0.99	0	-0.32
2	CB	Southern end	211	0.99	0	-0.26

(1) Modal class gap

The modal class gap (MCG), Δi , is the difference between two index numbers which denote the mode (most frequent) classes of the observed and estimated data histograms. It is represented as

$$\Delta i = i_{es_mode} - i_{ob_mode}, \quad (4.5)$$

where i_{ob_mode} and i_{es_mode} are the index numbers of the modes of the observed and estimated data histograms, respectively. This represents the accuracy of the position along the gate line where the largest number of vessels is observed.

(2) Modal value difference rate

The modal value difference rate (MVDR), r_{mv} , is the ratio of the difference between the mode values (amount of the most frequent class) of the observed and estimated histograms. It is represented as

$$r_{mv} = (es_hist(i_{es_mode}) - ob_hist(i_{ob_mode})) / ob_hist(i_{ob_mode}), \quad (4.6)$$

where $ob_hist(i_{ob_mode})$ and $es_hist(i_{es_mode})$ are the observed and estimated mode values, respectively. The rate represents the accuracy of the traffic volume at the most concentrated locations.

The fittings are performed on the data for the four paths of Period 1 at the northern and southern ends of the target area and near the restriction (central area). The open bar in Figure 7 show the estimated transverse distributions, which are generated from the estimated distribution function of path AC at the central area. Table 4 summarises the parameters obtained and the results of the goodness-of-fit test. The size of the histogram class was 2000.0 m. It can be observed from the table that the estimation of the location with high-density traffic is accurate, but the value is slightly underestimated.

According to the results in Section 4.2, only the traffic behaviour in the central area is affected by the variation in the situation. The traffic behaviours at the northern and southern ends of the area are not significantly affected. Therefore, it is necessary to predict transverse volume distribution in the central area. In the rest of the area, the observations are directly applicable.

Algorithm 1. Procedure for predicting transverse distribution

```

1: # initialise
2:  $\alpha \leftarrow \alpha_0$  # shape parameter
3:  $\beta \leftarrow \beta_0$  # scale parameter
4:  $\gamma \leftarrow \gamma_0$  # location parameter
5:  $x_{\text{area}} \leftarrow$  position # restricted area boundary position coordinate
6:  $u_0 \leftarrow$  Percentile( $f(x; \alpha_0, \beta_0, \gamma_0), 0.95$ ) # initial upper bound; 95 percentile value
7:  $l_0 \leftarrow$  Percentile( $f(x; \alpha_0, \beta_0, \gamma_0), 0.05$ ) # initial lower bound; 5 percentile value
8: # search new parameter set ( $\alpha_1, \beta_1, \gamma_1$ )
9:  $\alpha_1 \leftarrow \alpha_0$ 
10:  $u \leftarrow$  Percentile( $f(x; \alpha_1, \beta, \gamma), 0.95$ )
11:  $l \leftarrow$  Percentile( $f(x; \alpha_1, \beta, \gamma), 0.05$ )
12: while ( $|x_{\text{area}} - l| > th_1$ ) or ( $|u - u_0| > th_2$ ) : # comparing to thresholds
13:   while  $|x_{\text{area}} - l| > th_1$  :
14:      $\gamma \leftarrow \gamma + \Delta\gamma$ 
15:      $l \leftarrow$  Percentile( $f(x; \alpha_1, \beta, \gamma), 0.05$ )
16:    $\gamma_1 \leftarrow \gamma$ 
17:    $u \leftarrow$  Percentile( $f(x; \alpha_1, \beta, \gamma_1), 0.95$ )
18:   while  $|u - u_0| > th_2$  :
19:      $\beta \leftarrow \beta + \Delta\beta$ 
20:      $u \leftarrow$  Percentile( $f(x; \alpha_1, \beta, \gamma_1), 0.95$ )
21:    $\beta_1 \leftarrow \beta$ 

```

The traffic distribution is predicted based on the distribution function of the reference model. [Algorithm 1](#) describes the pseudocode of the procedure. First, the shape parameter, α , is determined by applying the values in the reference distribution, because there are almost no differences between the shape parameters of the two periods. The scale parameter, β , is adjusted while simultaneously adjusting the location parameter, γ . Here, γ is adjusted to reproduce the property of being concentrated outside the immediate vicinity of the restricted area, and β is adjusted to reproduce the property in which the distribution is unaffected at a distance.

A prediction model for the transverse distribution of Period 1 is created using the reference model, which is estimated from the data of Period 2. Each gate line is set in a direction orthogonal to the sailing direction of the ships, and individual tracks are predicted by connecting the passage positions on the gate lines. For the calculations, the restricted area is set 20 km from the power plant. It is also assumed that 5% of the passing ships crossed the eastern edge of the restricted area. (In the observation data, this percentage was 0%–12%, depending on the path).

The prediction results are evaluated in the same manner as the estimation results. [Table 5](#) compares the predicted distribution and original Period 1 data. From this table, the traffic volume at the concentration points tends to be slightly underestimated but the location of the concentration is predicted to be close. The transverse distribution of path AC, as an example, around the restricted area is indicated by the open bar in [Figure 8](#). The graph shows the general nature of the distribution is reproduced adequately.

The EF is calculated for each grid defined in [Section 4.1](#) using the tracks reproduced based on the previous estimation and prediction models. The resulting EF distributions are summarised in the maps shown in [Figure 9](#). [Figures 9\(a\)](#) and [9\(b\)](#) show the estimated and predicted results generated from the reference and prediction models, respectively. For comparison, [Figure 9\(c\)](#) shows the direct estimation from the observed data.

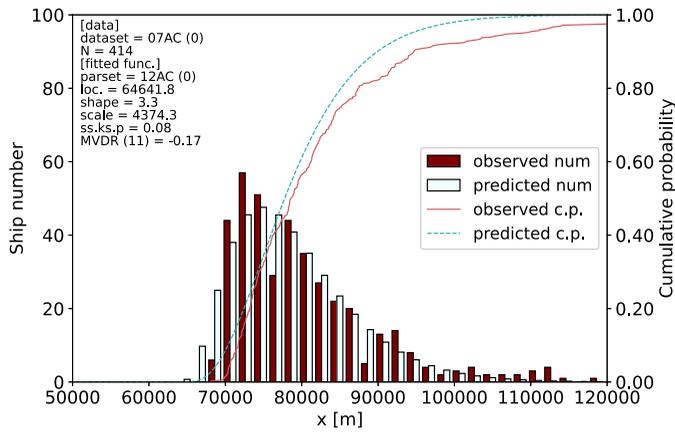


Figure 8. Observed and predicted density distributions of transverse position on path AC.

Table 5. Prediction results (prediction: period 1, reference: period 2).

Period	Path	Area	N	p-Val. (KS)	MCG	MVDR
1	AC	Central	414	0.08	1	-0.16
1	CA	Central	367	0.16	1	-0.39
1	BC	Central	220	0.00	-2	0.53
1	CB	Central	226	0.02	-1	0.30

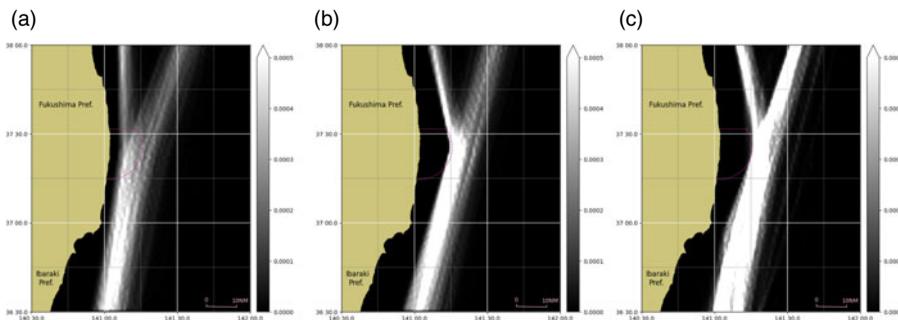


Figure 9. Distribution of EF calculated based on observed and predicted ship trajectories (times/s). (a) period 2 (estimated using the reference model), (b) period 1 (predicted using the prediction model), (c) period 1 (estimated directly from the observation data). Dashed lines represent the evacuation order zone of period 1.

It is observed that the high-encounter-frequency (HEF) locations are accurately predicted when the rule is applied. Specifically, the major HEF positions, i.e. the north branch, south end exit and central concentration positions outside the restricted area are accurately predicted. However, the frequency is slightly underestimated in some places. This is because traffic tends to be underestimated in areas with relatively high traffic volumes, as mentioned above.

4.4. Discussion

Sections 4.1–4.3 presented a systematic EF prediction method, EnFreq, based on the results of the impact analysis considering the introduction of restricted areas on traffic behaviour. This method calculates the EF by predicting the unknown traffic behaviour from the known traffic behaviour information. When a prediction model is created, the feature quantities and areas that must be predicted are reasonably narrowed down by sensitivity and behavioural change analyses.

This study considered a typical case in a coastal area for simplicity. Essentially, it is a coastal area with an open sea on one side, no large ports and only one navigational restriction. Additionally, class A AIS-equipped ships were targeted. Owing to these conditions, the main targets were merchant ships, and the size and speed ranges were limited. The demonstration showed that the influence of the traffic density on the EF formula was particularly high. In this instance, the actual increase in the travelling distance due to the detour was small, with a previous study suggesting an increase of approximately 10% in the speed (Yanagi et al., 2012). Such an environment is dominant in the waters near Japan; however, the effects vary depending on the conditions of the considered sea area and must be verified by sensitivity analysis. For instance, if a large detour is required, its effect on the ship speed must be analysed.

Furthermore, this calculation method can also be applied to ships observed by class B AIS and radar. However, in practice, the problem of insufficient information caused by the differences in the frequency of information transmission and in the information that can be acquired must be resolved, e.g. the size information of a ship cannot generally be acquired by radar.

This study demonstrated one of the most common scenarios in coastal areas, including encounters between two ships sailing in the opposite directions. For ships conflicting at other encounter angles, the calculation methods for various encounter angles, such as that proposed by Kawashima and Itoh (2019), or deterministic methods, such as that used in Miyake et al. (2017), should be considered.

The traffic in the target area is relatively simple, and the trajectories are grouped into two pairs of paths. Therefore, it is not very difficult to model the passage–position distribution of each path. For more complicated traffic flows, paths must be created by analysing OD pairs, considering the type and size of the ships or using the group generation techniques such as that proposed by Seshita et al. (2016).

A method is proposed to predict the density distribution by representing the transverse passage–position distribution with a three-parameter gamma distribution function. Using this function and adjusting the location and scale parameters, the changes in the passage–position distribution can be predicted. The function is suitable for expressing the density distribution that is commonly observed in coastal traffic, that is, asymmetrical and high kurtosis, but its applicability and the means of applying the function for increased complexity must be verified.

Accurate predictions of the location and degree of the high-traffic-density areas were essential for predicting the density distribution, and the metrics used to represent the accuracy were introduced. Using these metrics, it was confirmed that the predictions were made appropriately in the target sea area; however, the results may vary depending on the class width because the metrics use a histogram. Similarly, the encounter is calculated using the grid-based method, leading to a grid-size problem. Further consideration of these class widths and grid sizes is required to generalise the method.

The location and degree of the HEF can be obtained by the proposed method, which is useful for evaluating the proposed rules during design and when considering any amendments. Furthermore, Equation (2.1) predicts the number of collisions, which provides the basis for evaluating acceptability from a risk perspective.

5. Conclusion

In this study, a systematic method, called EnFreq, was proposed for the design of new rules in waters. It enables predicting the ship encounter frequency distribution in a sea area when ship traffic rules are changed. This method uses the observable traffic data and the restriction conditions to predict the EF after the introduction of new rules. Additionally, the concept of EF per unit area and unit time was introduced for understanding the occurrence of encounters.

The proposed method can be used in future traffic prediction as it is systematised to obtain results efficiently by effectively using the available data and limiting the elements that require difficult behaviour prediction as much as possible. In the demonstrations, the processes of narrowing down the items which require prediction, predicting the narrowed-down items and obtaining the resulting EF distribution were presented.

Based on the calculations, the limitations of the proposed method were analysed from several perspectives, such as issues related to the complexity of ship traffic, those related to the information acquired by class B AIS and radar, and those related to the application of this method to predict encounters involving ships from directions other than the opposite directions.

Ship encounters are the primary cause of future collisions, and to ensure the safety of marine traffic, it is important to highlight particularly hazardous locations, along with the degree of danger, when drafting new rules. In addition to the sea areas covered in this study and the initiatives introduced in Section 1, the proposed method can be applied to different sea areas and condition changes. This study will be extended to predict encounter frequencies under various settings.

Acknowledgements. The author is grateful to Dr. Rina Miyake and their colleagues from Risk Analysis Technology Research Group, National Maritime Research Institute, National Institute of Maritime, Port and Aviation Technology for valuable discussions. The author thanks Ms. Emiko Takanashi for her cooperation in treating traffic data.

Financial support. This work was partly supported by JSPS Grants-in-Aid for Scientific Research (KAKENHI) Grant number 18K04589.

References

- Altan, Y. C. and Otay, E. N. (2017). Maritime traffic analysis of the strait of Istanbul based on AIS data. *The Journal of Navigation*, **70**(6), 1367–1382.
- Blom, H., Bakker, B., Everdij, M. and Park, M. (2003). Collision Risk Modeling of air Traffic. *European Control Conference (ECC2003)*.
- Chan K. (2021). Collision Probability for A General Spacecraft, *31st AAS/AIAA Space Flight Mechanics Meeting*.
- Christensen, C. F., Andersen, L. W. and Pedersen, P. H. (2001). Ship Collision Risk for an Offshore Wind Farm. *Proceeding of the 8th International Conference on Structural Safety and Reliability*, Newport Beach, USA.
- Cucinotta, F., Guglielmino, E. and Sfravara, F. (2017). Frequency of ship collisions in the strait of Messina through regulatory and environmental constraints assessment. *The Journal of Navigation*, **70**, 1002–1022.
- DNV-GL (2015). Navigational risk assessment – Omø Syd offshore wind farm. Available at: https://ens.dk/sites/ens.dk/files/Vindenergi/navigational_risk.pdf (as of 2020.3.26)
- EMSA (2019). Annual overview of marine casualties and incidents 2019. Available at: <http://www.emsa.europa.eu/newsroom/latest-news/download/5854/3734/23.html> (as of 2021.5.9)
- Friis-Hansen, P. (2008). IWRAP MK II Working document Basic modelling principles for prediction of collision and grounding frequencies. Available at: https://www.iala-aism.org/wiki/iwrap/images/2/2b/IWRAP_Theory.pdf (as of 2020.3.26)
- Fujii, Y. (1983). Integrated study on marine traffic accidents. *IABSE Reports*, **42**, 91–98.
- Fujii, Y. and Tanaka, K. (1971). Traffic capacity. *The Journal of Navigation*, **24**(4), 543–552.
- Fujii, Y., Yamanouchi, H. and Matui, T. (1984). Survey on vessel traffic management systems and brief introduction to marine traffic studies. *Electronic Navigation Research Institute Papers*, **45**, 1–131.
- Fukushima Prefecture. (2019). Transition of evacuation designated zones. Available at: <http://www.pref.fukushima.lg.jp/site/portal-english/en03-08.html> (as of 2020.10.30)
- Goerlandt, F. and Kujala, P. (2014). On the reliability and validity of ship–ship collision risk analysis in light of different perspectives on risk. *Safety Science*, **62**, 348–365.
- Imazu, H. (2017). Evaluation method of collision risk by using true motion. *The International Journal of Marine Navigation and Safety of Sea Transportation*, **11**(1), 65–70.
- IMO. (2003). Guidance note on the preparation of proposals on ships' routing systems and ship reporting systems for submission to the sub-committee on safety of navigation. (IMO MSC/Circ.1060).
- IMO. (2016). Routing measures and mandatory ship reporting systems. Establishment of a recommended route off the western coast of Izu O Shima Island, Japan. (IMO NCSR 4/3.) Submitted by Japan.
- International Telecommunication Union (ITU). (2014). Technical characteristics for an automatic identification system using time division multiple access in the VHF maritime mobile frequency band (Recommendation ITU-R M. 1371-5).
- Itoh, H. and Miyake, R. (2019). Research on change of traffic safety accompanying the implementation of a new recommended route. *Proceeding of the 8th ICCGS*, Lisbon, Portugal, 247–254.
- Itoh, H. and Yakabe, F. (2014). Modeling ship traffic distributions in coastal areas. *Journal of the JASNAOE*, **19**, 235–244. (In Japanese).

- Itoh, H., Kaneko, F., Mitomo, N. and Tamura, K.** (2007). A Probabilistic Model for the Consequences of Collision Casualties. *Proceeding of the 4th ICCGS*, Hamburg, Germany, 201–206.
- Itoh, H., Ishimura, E., Yanagi, Y. and Mori, Y.** (2012). Cognitive Model of Maritime Navigation and Its Use for Collision Accident Analysis. *Proceeding of the 5th ICETET*, Himeji, Japan, 93–99.
- Japan Meteorological Agency (JMA).** (2019). Search past weather data. (Provisional translation) Available at: <https://www.data.jma.go.jp/obd/stats/etrn/index.php> (as of 2020.3.26) (In Japanese)
- JCG** (2012). JCG Annual Report. Available at: <https://www.kaiho.mlit.go.jp/info/books/report2012/html/top.html> (as of 2020.3.26) (In Japanese)
- JCG** (2018). Current status of marine accidents and countermeasures (Provisional translation). Available at https://www6.kaiho.mlit.go.jp/info/keihatsu/20180314_state_measure29.pdf (as of 2021.5.9) (In Japanese)
- Kawashima, S. and Itoh, H.** (2019). Assessment of ship encounter and collision in congested sea areas. *Proceeding of the 8th ICCGS*, Lisbon, Portugal. 267–274.
- Kawashima, S., Kawamura, Y., Itoh, H. and Fukuto, J.** (2015). Generation of Ship Traffic Flow Based on Principal Component Analysis of AIS Data and its Application to Ship Traffic Simulation for Evaluation of Encounter Probability. *Proceeding of Asia Navigation Conference 2015*, Kitakyushu, Japan. 255–264.
- Khaled, M. E. and Kawamura, Y.** (2015). Collision Risk Analysis of Chittagong Port in Bangladesh by Using Collision Frequency Calculation Models with Modified BBN Model. *Proceeding of the Twenty-fifth International Offshore and Polar Engineering Conference 2015*, Hawaii, USA.
- Kübler, H.** (1979). On the fitting of the three-parameter distributions Lognormal, Gamma, and Weibull. *Statistische Hefte*, **20**(2), 68–125.
- Lefèvre, S., Vasquez, D. and Laugier, C.** (2014). A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH Journal*, **1**(1), 1–14.
- Martins, M. R. and Maturana, M. C.** (2010). Human error contribution in collision and grounding of oil tankers. *Risk Analysis*, **30**(4), 674–698.
- Matui, T., Fujii, Y. and Yamanouchi, H.** (1983). Investigation on Marine Traffic in the Bisan Seto – No.1 Probability of Vessel Collision and Grounding –. *Electronic Navigation Research Institute Papers*, **43**. 1–19. (In Japanese).
- Miyake, R., Itoh, H., Nishizaki, C. and Fukuto, J.** (2016). Method of Safety Assessment for Establishing Ship Routeing System with Marine Traffic Simulation. *Proceeding of the 7th ICCGS*, Ulsan, Korea. 17–23.
- Miyake, R., Itoh, H., Nishizaki, C. and Fukuto, J.** (2017). Safety Assessment for Establishing Ships' Routeing – Recommended Route Off the Western Coast of Izu O Shima Island. *Proceeding of the 4th Asian Conference on Defence Technology*, Tokyo, Japan.
- Montewka, J., Hinz, T., Kujala, P. and Matusiak, J.** (2010). Probability modelling of vessel collisions. *Reliability Engineering and System Safety*, **95**, 573–589.
- Mujeeb-Ahmed, M. P., Seo, J. K. and Paik, J. K.** (2018). Probabilistic approach for collision risk analysis of powered vessel with offshore platforms. *Ocean Engineering*, **151**, 206–221.
- Nagatsuka, H. and Balakrishnan, N.** (2012). Parameter and quantile estimation for the three-parameter gamma distribution based on statistics invariant to unknown location. *Journal of Statistical Planning and Inference*, **142**, 2087–2102.
- National Institute of Standards and Technology (NIST).** (2012a). Kolmogorov-Smirnov Goodness-of-Fit Test. *Engineering Statistics Handbook*, 1.3.5.16. Available at: <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm>. (as of 2020.3.26)
- National Institute of Standards and Technology (NIST).** (2012b). Gamma Distribution. *Engineering Statistics Handbook*, 1.3.6.6.11. Available at: <http://www.itl.nist.gov/div898/handbook/eda/section3/eda366b.htm>. (as of 2020.3.26)
- Netjasov, F.** (2020). Conflict Risk Assessment Based Framework for Airspace Planning and Design, In book: Risk Assessment in Air Traffic Management, Chapter: 3, *IntechOpen*.
- Nyman, T.** (2009). Review of collision and grounding risk analysis methods which can utilize the historical AIS data and traffic patterns in seawaters. SKEMA Consolidation study SE3.2.1. Available at: <http://www.eutraproject.eu/uploadfiles/D2321-Review-of-collision-and-grounding-risk-analysis-VTT-03June09.pdf> (as of 2020.3.26)
- Pedersen, P. T.** (1995). Collision and grounding mechanics. *Proceeding of WEMT 95*, Copenhagen, Denmark, 125–157.
- Pedersen, P. T.** (2010). Review and application of ship collision and grounding analysis procedures. *Marine Structures*, **23**, 241–262.
- Pietrzykowski, Z., Wolejsza, P. and Magaj, J.** (2015). Navigators' behavior in traffic separation schemes. *The International Journal on Marine Navigation and Safety of Sea Transportation*, **9**(1), 121–126.
- Seshita, A., Kawamura, Y., Fukuto, J. and Itoh, H.** (2016). Proposal of Traffic Flow Tube Model for Collision Risk Assessment Method of Congested Sea Area. *Proceeding of Asia Navigation Conference 2016*, Yeosu, Republic of Korea, 212–221.
- Silveira, P. A. M., Teixeira, A. P. and Soares, C. G.** (2013). Use of AIS data to characterise marine traffic patterns and ship collision risk off the coast of Portugal. *The Journal of Navigation*, **66**, 879–898.
- Szlapczynski, R.** (2013). Evolutionary sets of safe ship trajectories within traffic separation. *The Journal of Navigation*, **66**, 65–81.
- Tsou, M.-C.** (2010). Discovering knowledge from AIS database for application in VTS. *The Journal of Navigation*, **63**, 449–469.
- Wang, Y., Monsere, C. M., Chen, C. and Wang, H.** (2018). Development of a crash risk-scoring tool for pedestrian and bicycle projects in Oregon. *Transportation Research Record*, **2672**(32), 30–39.

- Weng, J. and Xue, S.** (2015). Ship collision frequency estimation in port fairways: A case study. *The Journal of Navigation*, **68**, 602–618.
- Yanagi, Y., Itoh, H. and Mori, Y.** (2012). Effects on ship navigation arising from the caution zone. *The Journal of Japan Institute of Navigation*, **127**, 157–164. (In Japanese)
- Ylitalo, J.** (2009). Ship-Ship Collision Probability of the Crossing Area between Helsinki and Tallinn. *Helsinki University of Technology Systems Analysis Laboratory & Ship Laboratory Mat-2.4108 Independent research projects in applied mathematics*.
- Zou, Y., Zhang, H., Feng, D., Liu, H. and Zhong, G.** (2021). Fast collision detection for small unmanned aircraft systems in urban airspace. *IEEE Access*, **9**, 16630–16641.