Navigational safety and traffic pattern analysis using AIS data on the western coast of India

Saurav Mandal, Vishwanath Nagarajan* and Om Prakash Sha

Department of Ocean Engineering and Naval Architecture, Indian Institute of Technology Kharagpur, Kharagpur 721 302, India

Ship-borne automatic identification system (AIS) is required to be installed onboard ships as per the regulations of International Maritime Organization. The AIS data for Indian coast for three months was analysed to determine marine traffic safety, navigation pattern, etc. The entire western Indian coast was divided into 22 legs. The analysis utilized probit regression model to calculate quantitative collision risk values for identifying the risky zones in Indian coast. The methodology uses an estimate of the distance of closest point of approach and time for closest point of approach between any set of vessels in a zone to find the collision risk index.

Keywords: Automatic identification system, collision risk value, DCPA and TCPA, probit regression model.

RECENTLY for many seaports, navigational collisions are a major safety concern. Therefore, collision avoidance system (CAS) has come into focus in port navigation safety. The most widely used CAS on merchant ships is automatic radar plotting aid (ARPA). Previously, analysis of ship traffic was hindered due to lack of data. With the invention of automatic identification system (AIS), we have an abundance of marine navigation data. Using the newly available data, a general simulation model of marine traffic can be developed for worldwide use. In 2006, Hetherington *et al.*[1] showed that more than 90% of collisions occur due to human error. For this reason, researchers have focused on developing a probabilistic model of collision risk from the pilot perspective. Initially, collision risk values were calculated using traffic flow theory, according to which, collision accident of unit time of a certain area is defined as the risk at sea. Fujii and Tanaka stated that most navigators following the ships avoided entering the surrounding domain of the foregoing ships[2]. An elliptical-shape domain was initially proposed[2]. Later, based on the subjective concept of the navigator, a different kind of domain – one consisting of circles of different radii was presented[3]. Another method of evaluating collision risk (CR) values is by using distance of closest point of approach (DCPA) and time for closest point of approach (TCPA) as the proximity parameters. In the early stage, a simple model was constructed using DCPA, TCPA and their weightings[4]. With time, utilization of DCPA and TCPA in collision analysis increased noticeably. AIS data analysis was started recently. In a case study, speed distribution, course distribution, traffic density and spatial distribution of ships are presented for two different waterways – one in China and the other in the Netherlands[5]. The analysis was done by extracting information from AIS data. Previously it was difficult to calculate the CR value quantitatively. There were three different situations, viz. 'safe', 'potential collision risk' and 'direct collision risk'[6]. Later, a probabilistic model for collision risk was constructed which includes the perception of pilots as well as effects of environment[7]. We have collected AIS data from the Director General Lighthouse and Lightships (DGLL), National Data Centre (NDC), Mumbai to analyse the marine traffic safety, navigation pattern, etc. in Indian waters.

Previously, marine traffic surveys included continuous visual observation, radar observation and airborne photography for at least 72 h (ref. 8). The International Maritime Organization (IMO) has prepared guidelines for ships regarding installation of AIS[9]. Through marine traffic surveys and marine traffic engineering, we can compute the distribution of ship density[10], distribution of track[8], traffic flow, traffic volume, speed distribution[8], time pattern of traffic flow, ship domain[11], traffic capacity and many other navigational statuses. AIS messages are of two types, VDM (VHF-data link message) and VDO (VHF-data link own vessel report). The original information can be decoded after the decompression of VDM data. Automatic tracking system is used by ships and by vessel traffic services (VTS) to identify and locate ships by electronically exchanging data with other nearby ships and AIS base stations[12]. Different static and dynamic information can be obtained from AIS data: course over ground (COG), speed over ground (SOG), maritime mobile service identity (MMSI), etc.[13]. MMSI is a series of nine digits, which are sent in digital form over a radio frequency channel to uniquely identify ship stations, ship earth stations, coast stations, coast earth stations, and group calls[13]. The AIS transponder works in continuous mode regardless of whether the ship is offshore, in coastal or inland waters, or at anchor. Moored and anchored ships broadcast their positions less frequently.

*For correspondence. (e-mail: vishwanath_n@naval.iitkgp.ernet.in)

Figure 1. Legs of western Indian coast.

AIS data analyses

We collected AIS data from November 2012 to January 2013, because the traffic density of western coast is more than that of the eastern coast (Indian Port Association, we confined our study to the western Indian coast). To perform the analysis, we divided the western Indian coast into 22 number of legs. Of these, 11 are within the exclusive economic zone (EEZ) and the remaining to the territorial sea (Figure 1). The legs named 'CW' are within EEZ and the legs indicated by OW belong to the territorial sea. Every corner point of the legs is represented by alphanumeric characters such as a1, a2, b1, b2, c1, etc. The latitude and longitude of every point are shown in Table 1.

In order to carry out zone (leg)-wise risk assessment, it is important that meaningful data are extracted from the available lot for individual legs. MMSI no. 1193046 is the most common invalid MMSI that occurs in transmitting the ship information. SOG with value 102.3 knots is the default raw value for SOG assigned to AIS unit when the speed info is not received. If true Heading is 511, it is also considered as noise. For our analysis, vessels with SOG less than 0.2 knots are also not included. Because, it is the vessels in the traffic waterways, which contribute to the collision risk in the shipping lanes and not the static vessels, which are either anchored, moored or lie at the fairways of the ports. The final dataset for analysis includes ship information, time, MMSI, latitude, longitude, SOG and COG. For each leg, the datasets are further grouped on the basis of MMSI number. Thus, all the information of a particular ship from the time it enters and exits the leg is recorded.

First, we have to identify the ships, which are within the leg of our consideration. We assume the coordinates of the ship is (x, y). If the coordinates of the legs are (x_1, y_1), (x_2, y_2), (x_3, y_3) and (x_4, y_4), the ship has to satisfy eqs (1) to (4) to be in that leg

$$y - y_1 \geq \frac{y_2 - y_1}{x_2 - x_1}(x - x_1), \tag{1}$$

$$y - y_2 \leq \frac{y_3 - y_2}{x_3 - x_2}(x - x_2), \tag{2}$$

$$y - y_3 \le \frac{y_4 - y_3}{x_4 - x_3}(x - x_3), \tag{3}$$

$$y - y_4 \ge \frac{y_1 - y_4}{x_1 - x_4}(x - x_4). \tag{4}$$

Figure 2 shows one such leg with coordinates as stated above. The two proximity indicators used in the risk calculation are DCPA and TCPA. These two parameters represent spatial and temporal closeness between a pair of ships at any instant of time respectively. In Figure 2, DCPA and TCPA are explained showing a case of vessel interaction. DCPA and TCPA between two vessels can be derived using eqs (5) to (8)

$$\mathrm{DCPA} = D \times \cos(C_{rt} - B_{rt}), \tag{5}$$

$$\mathrm{TCPA} = \frac{D \times \sin(C_{rt} - B_{rt})}{V_{rt}}, \tag{6}$$

$$V_{rt} = \sqrt{V_0^2 + V_t^2 - 2V_0 V_t \cos(c_t - c_0)}, \tag{7}$$

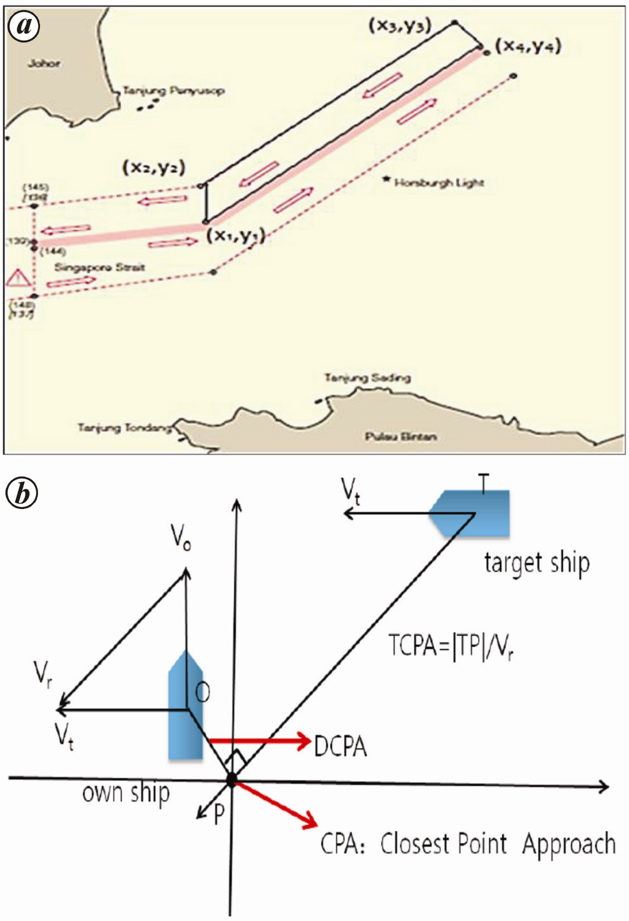

Figure 2. ***a***, Sample leg with coordinates (x_1, y_1), (x_2, y_2), (x_3, y_3) and (x_4, y_4). ***b***, DCPA and TCPA between two vessels.

$$C_{rt} = \cos^{-1}\left(\frac{V_0 - V_t \cos(c_t - c_0)}{V_{rt}}\right), \tag{8}$$

where D, C_{rt}, B_{rt} and V_{rt} are the absolute distance, relative course, relative bearing and relative velocity respectively. True bearing is measured with respect to true north in clockwise direction. V_0 and V_t are the velocity of own ship (with reference to which DCPA and TCPA are measured) and the target ship respectively. Similarly c_0 and c_t are the courses of own ship and target ship respectively. DCPA, TCPA and V_{rt} are measured in cable lengths, minutes and m/s respectively. Figure 3 shows a flow diagram for calculating the collision risk.

A probabilistic model of collision risk is used by formulating an ordered probit regression model and calibrated using perceived risk data. The risk of collision in any vessel interaction mainly depends on its size. Research shows that the ease of speed adjustment and maneuverability decreases as vessel sizes increase. For this reason, different vessel sizes have different collision

Table 1. Coordinates of the corner points of the legs

Points	Latitude	Longitude
a1	20°42′6.6″N	70°48′19.4″E
a2	21°2′37.8″N	72°31′22.9″E
a3	21°2′37.8″N	72°43′22.9″E
a4	19°29′23.5″N	69°12′53″E
a5	19°29′23.5″N	72°32′41″E
a6	19°29′23.5″N	72°44′41″E
b1	18°29′25.3″N	69°22′14.2″E
b2	18°29′25.3″N	72°42′02″E
b3	18°29′25.3″N	72°54′02″E
c1	15°56′13.5″N	70°00′17.1″E
c2	15°56′13.5″N	73°20′05.1″E
c3	15°56′13.5″N	73°32′05.1″E
d1	14°53′47.3″N	70°33′36.2″E
d2	14°53′47.3″N	73°53′24.2″E
d3	14°53′47.3″N	73°65′24.2″E
e1	13°40′46.4″N	71°07′53.5″E
e2	13°40′46.4″N	74°27′41.5″E
e3	13°40′46.4″N	74°39′41.5″E
f1	11°52′04″N	71°51′07.6″E
f2	11°52′04″N	75°10′55.6″E
f3	11°52′04″N	75°22′55.6″E
g1	11°09′40.5″N	72°16′20.8″E
g2	11°09′40.5″N	75°36′08.8″E
g3	11°09′40.5″N	75°48′08.8″E
h1	10°08′04.2″N	72°39′03.6″E
h2	10°08′04.2″N	75°58′51.6″E
h3	10°08′04.2″N	75°70′51.6″E
i1	9°42′10.9″N	72°45′28.9″E
i2	9°42′10.9″N	76°05′16.9″E
i3	9°42′10.9″N	76°17′16.9″E
j1	8°38′54.6″N	73°14′38.5″E
j2	8°38′54.6″N	76°34′26.5″E
j3	8°38′54.6″N	76°46′26.5″E
k1	8°04′41.2″N	74°01′13.6″E
k2	8°04′41.2″N	77°21′01.6″E
k3	8°04′41.2″N	77°33′01.6″E

risks. In order to calculate the perceived collision risk, vessels are categorized into four groups (Table 2). In our approach, we directly study vessels' real time dynamics measured by AIS. The vessel hydrodynamics response and control characteristics are already embedded in the experiment data. The risk intensity is categorized into five levels namely:

Level 1: very high, level 2: high, level 3: moderate, level 4: low, level 5: safe (Chin *et al.*[7]). Equation (9) explains the ordered probit model, which is the main tool to calculate the collision risk value in this study

$$y_i^* = \beta \boldsymbol{X}_i + \varepsilon_i, \tag{9}$$

where y_i^* is a continuous latent variable measuring collision risk for the i^{th} set of $\boldsymbol{X}$. $\boldsymbol{X}_i$ is a vector of the independent variables (DCPA and TCPA measured between own vessel and target vessel). β is a vector of regression coefficients. Coefficient ε_i is a normally distributed random error term with 0 mean and 1 as standard deviation. y_i^* is mapped on to an observed ordinal variable y by

$$y_i = m, \text{ if } \tau_{m-1} \le y_i^* < \tau_m; \text{ for } m = 1 \text{ to } J,$$

where J is the ordinal categories in y. Coefficient τ describes the boundaries of risk levels.

Here $m = 1$ implies very high risk and $m = 5$ is safe. The way in which the latent variable is connected to the discrete risk levels is explained in eq. (10)

$$\begin{aligned}
y_i &= 1 \text{ if } \infty \le y^* < \tau_1 \text{ (VHR)},\\
y_i &= 2 \text{ if } \tau_1 \le y^* < \tau_2 \text{ (HR)},\\
y_i &= 3 \text{ if } \tau_2 \le y^* < \tau_3 \text{ (MR)},\\
y_i &= 4 \text{ if } \tau_3 \le y^* < \tau_4 \text{ (LR)},\\
y_i &= 5 \text{ if } \tau_4 \le y^* < \infty \text{ (Safe)},
\end{aligned} \tag{10}$$

where VHR: very high risk, HR: high risk, MR: moderate risk, LR: low risk. The probability of risk level m for given X_i can be predicted by eq. (11).

$$\hat{\Pr}(y = m|\boldsymbol{X}_i) = F(\hat{\tau}_m - \hat{\beta}\boldsymbol{X}_i) - F(\hat{\tau}_{m-1} - \hat{\beta}\boldsymbol{X}_i), \tag{11}$$

where F is the cumulative distributive function of ε.

Here

$$\sum_{m-1}^{j} \hat{\Pr}(y = m/X_i) = 1.$$

After predicting the probabilities of each risk level, the associated collision risks can be computed using RS_m. RS_m is the risk scores assigned to every risk level. The values of RS_m were defined based on the thresholds. RS_m is actually the probability of collision risk for risk level m. Risk scores for VHR and safe level are 1 and 0 respectively. The collision risk for the given X_i (a set of DCPA and TCPA values between the own vessel and target vessel) is computed as shown in eq. (12)

$$C(X_i) = \sum_{m=1}^{j} RS_m \times \hat{\Pr}(y = m \mid X_i). \tag{12}$$

In this article $C(X_i)$ is considered as the collision risk (CR) value. For our study, the regression coefficients are taken from Chin *et al.*[7]. The values of regression coefficients (Table 3) vary with the gross tonnage of ships and visibility conditions. Visibility is divided into two categories: 'day time' and 'night time'. Following this procedure,

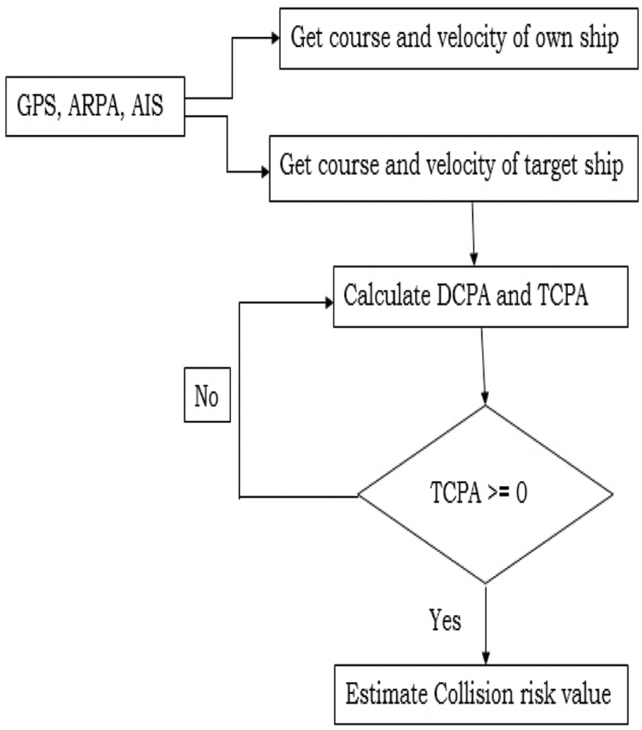

Figure 3. Flow diagram to estimate CR values.

Table 2. Vessel categories and their descriptions

Vessel categories	Description
VC1	If 300 ≤ GT ≤ 12,000
VC2	If 12,000 < GT ≤ 20,000
VC3	If 20,000 < GT ≤ 75,000
VC4	If 75,000 < GT

Table 3. Regression coefficients[7]

	VC1		VC2		VC3		VC4	
	Day	Night	Day	Night	Day	Night	Day	Night
Coefficients								
$\beta(1)$	0.266	0.2179	0.5611	0.6502	0.2641	0.271	0.2641	0.271
$\beta(2)$	0.1168	0.0902	0.3278	0.2637	0.1151	0.1181	0.1151	0.1181
Thresholds								
τ_1	0.2716	0.3271	0.7505	1.3021	0.3212	0.5363	0.3212	0.5363
τ_2	1.0468	1.2946	2.5342	3.3943	1.5432	1.8126	1.5432	1.8126
τ_3	2.1088	1.9947	4.6098	5.9758	2.3581	2.7565	2.3581	2.7565
τ_4	3.1519	3.0112	6.9348	8.5806	3.4408	3.9437	3.4408	3.9437
Risk scores								
RS_{VHR}	1	1	1	1	1	1	1	1
RS_{HR}	0.9138	0.8913	0.8917	0.8482	0.9066	0.8640	0.9066	0.8640
RS_{MR}	0.6678	0.5700	0.6345	0.6044	0.5515	0.5403	0.5515	0.5403
RS_{LR}	0.3309	0.3375	0.3352	0.3035	0.3146	0.3010	0.3146	0.3010
RS_{safe}	0	0	0	0	0	0	0	0

collision risk matrix is generated at each instant of time, as shown in eq. (13)

$$\mathrm{CR}(t) = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix}, \tag{13}$$

where $C_{ij} = 0$, for $i = j$ or TCPA < 0; $C_{ij} = C$, otherwise. $0 < C < 1$.

Now $\mathrm{CR}_{\max} = \mathrm{Max}\,(C_{ij})$; $i, j = 1, 2, \ldots, n$. Hence, a time series of all CR values are generated on hourly basis.

The ordered probit model, used in our analysis distinguishes different risk levels by developing a set of thresholds. The threshold values of τ, $\hat{\beta}_1$ and $\hat{\beta}_2$ are provided in Table 3. Since these values are on the latent variable case, the structural model can be modified as eq. (14)

$$\hat{\tau}_m = \hat{\beta}_1 \times \mathrm{DCPA} + \hat{\beta}_2 \times \mathrm{TCPA};\ \text{for } m = 1 \text{ to } j-1. \tag{14}$$

From this equation, for each risk level, two boundary points on DCPA and TCPA can be obtained. Putting TCPA = 0 in the equation yields the value of $\mathrm{DCPA}_{\mathrm{TCPA}=0}$. This value is known as distance safety margin (DSM). For each risk level, pilots would allow DSM as the maximum passing distance between two ships. Similarly $\mathrm{TCPA}_{\mathrm{DCPA}=0}$ is known as the time safety margin (TSM) and this depicts the marginal time before a collision. CR value corresponding to DSM and TSM can be considered safe for own vessel and target vessel. So if the calculated value is greater than the safe CR value, the situation may be risky. In this way, CR value is related to risk factors. In our analysis, we have calculated CR values with a 10 s time interval. Therefore, in an hour CR values are generated 361 times. Since it is impossible to get updates from all ships every each 10 s, we used linear interpolation method. For every leg, we considered all values of the CR matrix and plotted the probability density function (PDF) of all the values. In this case, we combined the CW leg with their corresponding OW leg but considered them as one leg. For example, leg 1 CW and leg 1 OW are combined and considered as leg 1. We derived a number of non-zero CR values for each leg. This will help us identify the most risky leg. The number of north bound and south bound ships for the whole western coast has also been calculated separately for three different months to understand the traffic pattern of western Indian coast.

In our CR analysis, we categorized the vessels into four types based on their gross tonnage (GT)[7]. We considered all ships of western Indian coast for every month and constructed a histogram based on their GT. Using marine traffic simulation system the difficulty of sailing a vessel can be solved up to a certain limit[14]. This simulation system uses fuzzy reasoning to predict collision risk using DCPA and TCPA. The ship domain theory was also used to find out the critical CR value[14]. According to this study, the value of CR in a near miss situation is 0.7. We have considered 0.7 as the critical CR value in our study as an initial approach towards our analysis. We plan to find out the exact CR value which will be suitable for our process of analysis in future.

Statistical analysis on the AIS data can also be carried out to obtain detailed information regarding navigation. An example of such an analysis is the determination of bivariate probability distribution or joint probability distribution between different variables. Examples of such pair of variables are (DCPA, TCPA), (speed dispersion, CR), (COG dispersion, CR), etc. In our analysis, the pair of variables are DCPA and TCPA. The number of ships within the variable range divided by the total number of ships yields the probability of each distribution cell. In addition, the dependency between the variables can be obtained by calculating the covariance between them. If the critical values of DCPA and TCPA considered risky for a pair of ships are known, the number of ships facing a risk can be calculated from the probability distribution.

Results and discussion

The PDF of CR values have been plotted for the months November 2012, December 2012 and January 2013 (Figures 4–6). From the analysed data, we observe that the CR values for this marine traffic zone mainly lie between zero and 0.1. It can be concluded that from the aspect of marine traffic density, the investigated traffic legs are safe. We have calculated DSM and TSM values for vessel class 3. Most ships in the waterway belong to this class. So we confined our calculation to this class. The day time was considered as navigating time, because it is safer as compared to the night time. We took threshold value (τ_m) for safety which means $m = 4$. Any value of DCPA and TCPA greater than DSM and TSM is ideally considered safe. The collision risk value corresponding to these DSM and TSM values is found to be 0.25 and 0.22 (approximately) respectively. In the first case we took DCPA = DSM and TCPA = 0. In the second case we took DCPA = 0 and TCPA = TSM. Both calculated values are greater than 0.1. Therefore, CR values less than 0.1 can be treated as safe. However, in the case of legs 2, 7 and 9, slight variation in the pattern can be observed. In these legs most CR values lie between 0 and 0.1. However, higher CR values are also probable. For example in leg 2, there is a marginal probability of getting values between 0.5 and 0.6 for November. In December, CR values can be observed between 0.2 and 0.8 in this leg. Similarly, in leg 7, there is a marginal probability that CR values can be between 0.7 and 0.8. In case of leg 9, CR values between 0.2 and 0.9 have a marginally higher probability. This phenomenon is true for all the three months. While analysing the AIS data for January, it was observed that the CR values between 0.3 and 0.5 of leg 4 have marginally higher probability. The geographical locations of some of the busiest ports in the Western Coast of India are Kandla (23°01′N, 70°13′E), the twin ports of Mumbai (18°54′N, 72°49′E) and JNPT (18°56.43′N, 72°56.24′E),

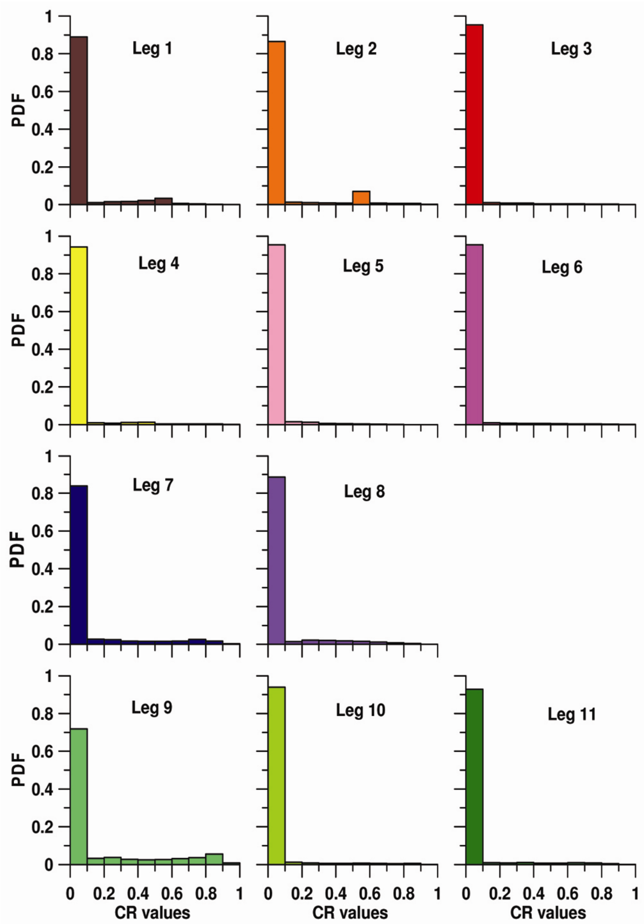

Figure 4. Variation of PDF of CR values of different legs (November 2012).

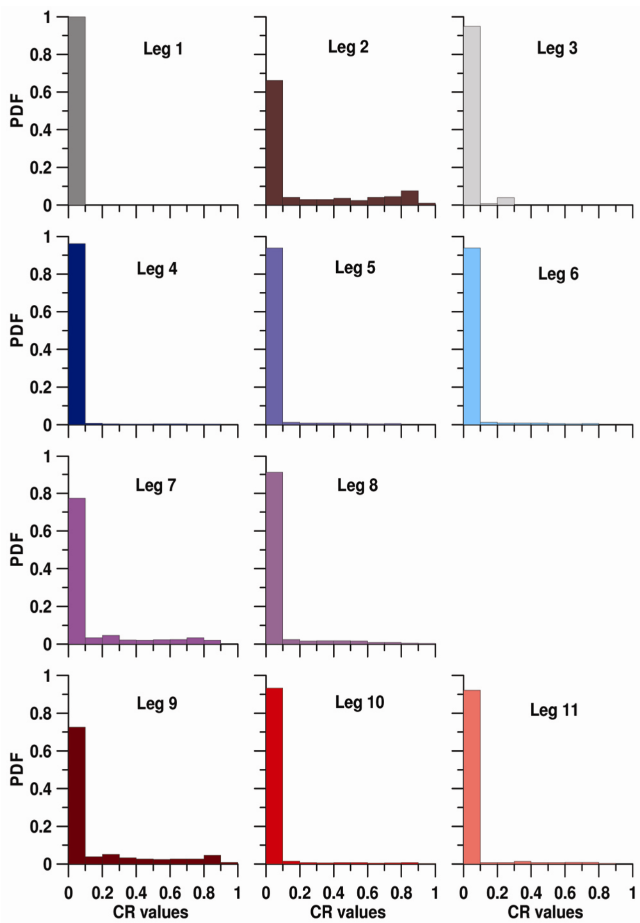

Figure 5. Variation of PDF of CR values of different legs (December 2012).

Mormugao (15°25′N, 73°48′E), Mangalore (12°55′N, 74°48′E) and Cochin (9°58′N, 76°16′E). It is observed that Mumbai and JNPT ports are near leg 2, Mangalore port is near leg 7 and Cochin port is near leg 9. Due to these busy ports, the number of ships in these legs may be higher as compared to others. As per the statistics[15], Kandla port is one of the busiest Western Indian coast port. However, leg 1 falls much below the geographical area of Kandla port. Besides the traffic density, the two-way navigation pattern also partly contributes to higher CR values. This analysis is useful in identifying the commercially active ports in a leg. In Figure 7, the number of south-bound and north-bound ships for each of the three months is shown. It is observed that numbers of south-bound ships are consistently higher as compared to northbound ships. This analysis has no direct connection to our CR value analysis. Both types of ships (south-bound and north-bound) normally follow different lanes for their navigation. In some waterways there are traffic separation schemes. Such waterways are safer as the chance of head-on collision is less. If there is a chance of head-on collision, that will also reflect in our analysis. In such cases, the value of DCPA and TCPA will decrease abruptly as the relative velocity is very high. This analysis was mainly done to get an overall view of the traffic pattern of Indian Western coast. There is a trade pattern for oil tankers. The south-bound ships are expected to be laden as they arrive from Persian Gulf area with cargo. However, north-bound oil tankers are expected to be in

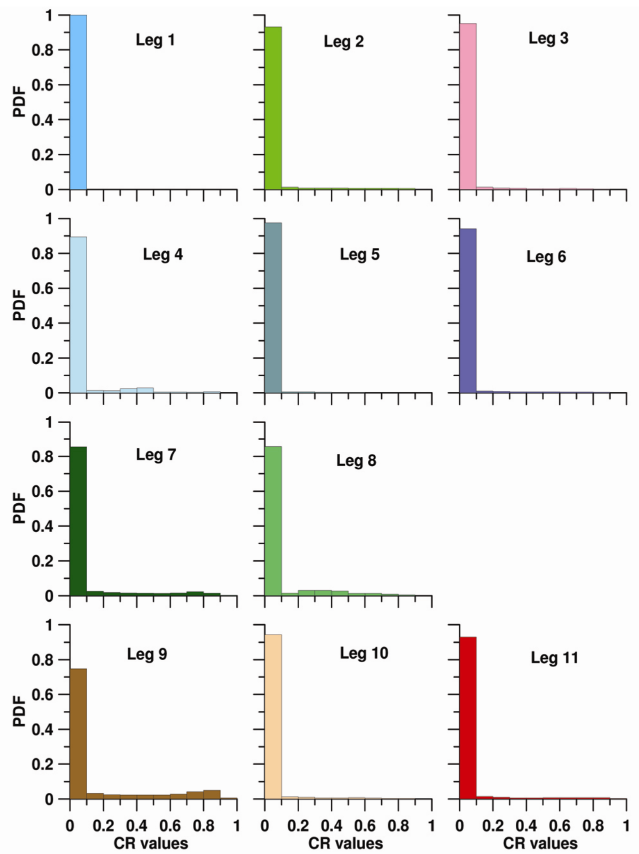

Figure 6. Variation of PDF of CR values of different legs (January 2013).

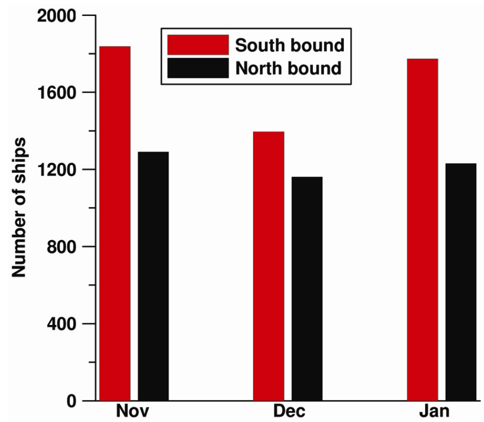

Figure 7. Variation of the number of north-bound and south-bound ships for November 2012, December 2012 and January 2013.

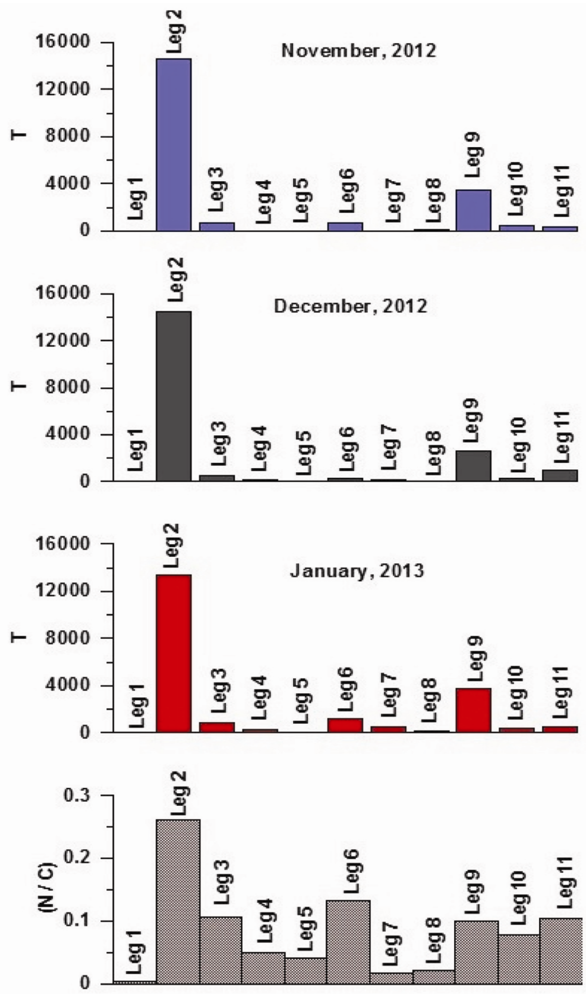

Figure 8. (Top 3) Number of occurrences of CR values higher than 0.7. (Bottom 1): Proportion of non-zero CR values for each leg. The analysis is for the months November 2012, December 2012 and January 2013.

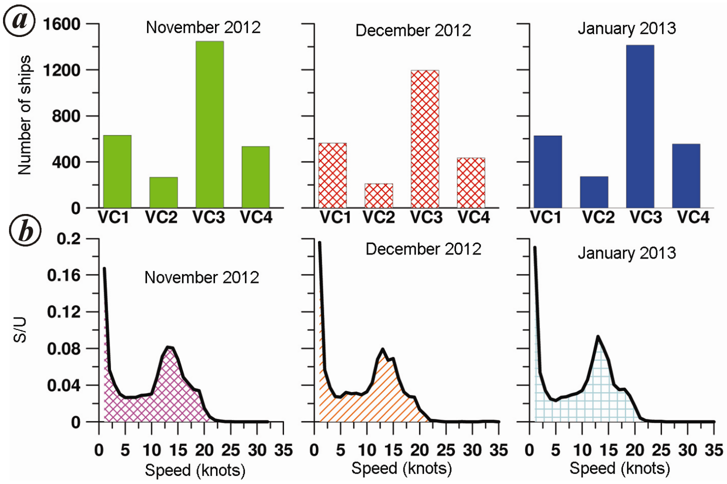

Figure 9. ***a***, Number of ships belonging to different tonnage categories. ***b***, Speed distribution of the ships for three different months.

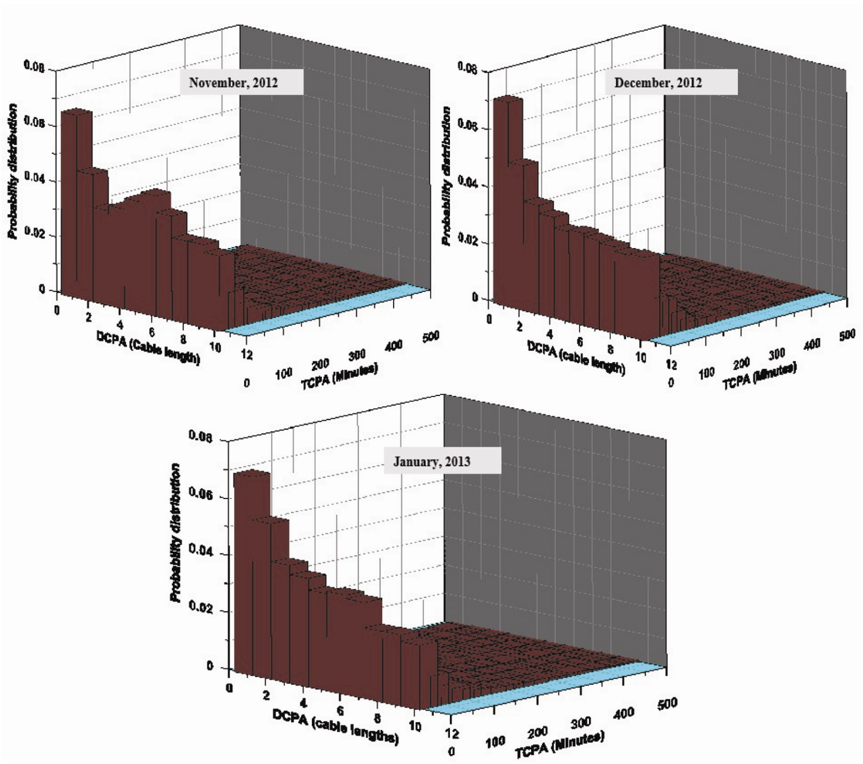

Figure 10. Bivariate frequency distribution for leg 2. Variables are DCPA and TCPA.

ballast draft condition. However, for other cargo ships no such trends could be expected. Presently we do not distinguish the ship's loading condition. However, this will have lot of bearing on the ship's collision response analysis. These aspects will be studied in our future work. For each leg, we calculated the number of times CR values exceeded 0.7 (Figure 8). T represents the number of times CR value has exceeded the critical CR value 0.7. In this analysis also, leg 2 appears as the most critical zone from the aspect of navigation safety. In the month of November, CR values in leg 2 exceeded the critical value about 14,000 times. This number is 14,000 and 13,000 for December and January respectively. Besides leg 2, CR values of leg 9 exceed the critical CR values for maximum number of times. In Figure 8, the number of non-zero CR values for each leg has been presented. The values are normalized with the total number of cases for that zone. In the figure, N represents the number of non-zero CR values and C represents the total number of CR values. The analysis was done for all the three months together. The maximum number of non-zero CR values is found in leg 2. Legs 6 and 9 also have a noticeable number of non-zero CR values. We calculated the CR values after categorizing the ships based on their gross tonnage (GT). It is observed that for each one of the three month investigated, the maximum number of ships belongs to the VC 3 category. For this category, the GT is in the range of 20,000–75,000. For most of the commercial merchant ships, GT is in the VC 3 category (L_{bp} 225–275 m)[16]. The above analysis is presented in Figure 9. We also calculated the speed distribution of ships for the three months separately. This analysis has been done for the entire western coast of India. First, we determined the maximum speed of ship for each month. Based on the recorded data, we fixed the speed limit for analysis as 35 knots as higher values for merchant ships are unlikely. We considered the incremental speed-range of 1 knot and calculated the frequencies (Figure 9). S represents the calculated frequency, and U represents the total number of AIS updates for the whole month. AIS data generally updates depending on the navigating speed and manoeuvering condition. Therefore, the recorded AIS data for a zone will have repeated updates from the same ship and our calculation of the number of occurrences will have repetitive data. We divided the calculated frequencies by the total number of AIS updates to normalize the data. It can be observed that the number of stationary ships waiting in the anchorage is relatively high. Of the navigating ships, most of them steam at typical speeds for commercial merchant ships (10 to 20 knots). The bivariate discrete frequency distribution analysis was done based on two variables, DCPA and TCPA. The unit of DCPA is one cable length (1 nautical mile = 10 cable length) while the unit of TCPA is 'minute'. After observing the analysis, we fixed our maximum TCPA as 500 minutes and the maximum DCPA as 10 cable lengths as the frequency is maximum in this range (TCPA for each cell is 25 min DCPA is 1 cable length). The bivariate frequency distribution is shown in Figure 10. Cells with maximum frequencies are within the TCPA range of 75 min and DCPA range of 5 cable lengths. This phenomenon is almost the same for three months. The data analyses for three months have been carried out separately.

Conclusions

Based on our research and analysis we can draw the following conclusions:

(1) From the PDF of CR values, it has been observed that in most cases, CR values lie between 0 and 0.1. Therefore, it can be stated that the situation is safe enough. A slightly different pattern was observed in legs 2, 7 and 9. The CR values in these zones are little higher compared to the other legs. Now, the twin ports of Mumbai and JNPT are in the vicinity of leg 2, Mangalore port is in the vicinity of leg 7 and Cochin port is in the vicinity of leg 9. Presence of these busy ports has increased the number of ships in these legs. This phenomenon can be considered as the reason for high CR values in these three legs.

(2) The number of north-bound and south-bound ships was also calculated for November–January. For each month, it was observed that the number of north-bound ships was less compared to south-bound ships. This gives an apparent idea of the traffic pattern of western Indian coast.

(3) It was observed that each month, the number of type three ships (GT range: 20,000–75,000) was greatest for the Western Indian Coast.

(4) Maximum number of ships was found to be steaming within a speed range of 10–20 knots. This speed range is the design speed of commercial merchant ships.

1. Hetherington, C., Flin, R. and Mearns, K., Safety in shipping: the human element. *J. Safety Res*., 2006, **37**, 401–411.
2. Fujii, Y., Yamanouchi, H., Tanaka, K., Yamada, K., Okuyama, Y. and Hirano, S., The behavior of ships in limited waters. *Electron. Navigation Res. Inst. Pap.*, 1978, **19**, 1–14.
3. Godwin, E. A., Statistical study of ship domain. *J. Navigation*, 1975, **19**, 1–14.
4. Kearon, J., Computer programs for collision avoidance and traffic keeping. In Conference on Mathematical Aspects on Marine Traffic, Academic Press, London, 1977.
5. Xiao, F., Lighter, H., Gulijk, C. V. and Ale, B., Comparison study on AIS data of ship traffic behavior. *Ocean Eng.*, 2015, **95**, 84–93.
6. Pedersen, P. T., Collision and grounding mechanics. *Danish Soc. Naval Arch. Mar. Eng.*, 1995, **125**, 57.
7. Chin, H. C. and Debnath, A. K., Modelling perceived collision risk in port water navigation. *Safety Sci.*, 2009, **47**(10), 1410–1416.
8. Yao, C. and Liu, Z., Distribution diagram of ship tracks based on radar observation in marine traffic survey. *J. Navigation*, 2010, **63**(1), 129–136.
9. Solas, I. M. O., *International Convention for the Safety of Life at Sea*, International Maritime Organization, London, 2003.
10. Beattie, J. H., Encounter rates in marine traffic separation schemes. *J. Navigation*, 1973, **24**(3), 325–340.
11. Celletti, M. D., Traffic models for use in vessel traffic system. *J. Navigation*, 1978, **31**(3), 104–116.
12. Tsou, M. C., Discovering knowledge from AIS database for application in VTS. *J. Navigation*, 2010, **63**, 449–469.
13. Mandal, S., Nagarajan, V. and Sha, O. P., AIS data, its application and analysis. In Proceedings of International Conference on Computing in Mechanical Engineering, Kochi, India, 2015.
14. Hasegawa, K. and Yamazaki, M., Qualitative and quantitative analysis of congested marine traffic environment – an application using marine traffic simulation system. *Int. J. Mar. Navigation Safety Sea Transport*., 2003, **7(2)**, 179–184.
15. Indian Ports Association; http://www.ipa.nic.in (accessed on 12 April 2016).
16. *Japanese Handbook of Ship Design*, JASNAOE, Japan, 2012.

ACKNOWLEDGEMENTS. This study is partially sponsored by the Ministry of Shipping, Government of India under the grant no. SY-13014/16/2010-SBR dated 13.06.2011. The authors thank Indian Maritime University, Visakhapatnam, India for their collaboration in this work.

Received 27 September 2016; revised accepted 21 February 2018

doi: 10.18520/cs/v114/i12/2473-2481