A Statistical Model for Vessel-to-Vessel Distances to Evaluate Radar Interference

Gaspare Galati, Gabriele Pavan, Francesco De Palo and Giuseppe Ragonesi

(*Department of Electronic Engineering, University of Rome Tor Vergata – Italy*)
(E-mail: gabriele.pavan@uniroma2.it)

Maritime traffic has significantly increased in recent decades due to its advantageous costs, delivery rate and environmental compatibility. With the advent of the new generation of marine radars, based on the solid-state transmitter technology that calls for much longer transmitted pulses, the interference problem can become critical. Knowing the positions and the heights of the ships, the mean number of the vessels in radar range can be estimated to evaluate the effects of their mutual radar interferences. This paper aims to estimate the probability density function of the mutual distances. The truncation of the density function within a limited area related to horizon visibility leads to a simple single-parameter expression, useful to classify the ships as either randomly distributed or following a defined route. Practical results have been obtained using Automatic Identification System (AIS) data provided by the Italian Coast Guard in the Mediterranean Sea.

1. INTRODUCTION.   Maritime traffic levels are related to economic growth: the international shipping industry is responsible for delivering about 90% of all trade worldwide by volume (with 7 to 9 billion tons loaded per year), and it is vital for bulk transport of raw materials, oil and gas. The linear regression between the economic growth of the nations in the Organisation for Economic Cooperation and Development (OECD) shows a 4% increase of imports and exports for a 1% increase in Gross Domestic Product (GDP). So, marine transportation is an integral, although sometimes less visible, part of the global economy. Other important marine activities include passenger transportation (ferries and cruise ships), national defence and fishing. If we also consider pleasure boats, even forgetting the millions of smaller leisure boats worldwide, the spatial density of marine traffic increases significantly, especially in areas close to harbours and coasts.

To improve the safety and the efficiency of marine traffic, as well as to protect the environment, Vessel Traffic Service (VTS) definitions and regulations were introduced by the International Maritime Organization (IMO) in 1985 and then updated in 1997 with the Resolution A.857(20) (IMO, 1997). VTS makes use, among other sensors, of powerful

coastal radars to obtain an up-to-date marine traffic image. Vessels themselves can carry radars for navigation purposes. For decades these navigation radars have been designed as cheap pulsed magnetron devices emitting very short (e.g. less than $10^{-1} \mu$ s) pulses with high peak power of the order of tens of kW. More recently, new solid-state technology has entered marine radar design (Zhang et al., 2013; Amato et al., 2010; Nelander and Tòth-Pàl, 2009). With a solid-state power amplifier, marine radars are becoming more and more sophisticated and agile in frequency as well as in waveform in order to get better performance by pulse compression, all features that were not possible with the old magnetron transmitter. Since the solid-state amplifiers' power is of the order of tens of W, significantly less than magnetrons', they have to transmit longer pulses (e.g. up to 90–100 μ s) in order to achieve the same energy on target, i.e. keep the detection probability unchanged with respect to magnetron radars. Their longer pulses cause a time-domain overlapping (interference) of several returns belonging to different radars, that is to different vessels. The main factors that affect interference between radars are: frequency, pulse rate, rotation rate and distance. The time-domain asynchronous interference in the older magnetron radar can be treated using simple Interference Rejection (IR) techniques. With the "long pulse" of solid state radars, the interference problem is much more marked and calls for a more complex interference mitigation technique (Galati et al., 2015). The various above-mentioned factors are included in a mathematical model which is fully described in Galati et al. (2016).

In this paper we focus on the distance between radars. Using real-world AIS data for the Mediterranean Sea, a statistical model of the mutual distances between vessels has been evaluated. By this it is possible to estimate the mean number of nearby ships in radar range, where the visibility depends on the radar antennae's heights.

The problem of knowing the statistical mutual distance distribution between vessels has not been addressed in literature so far. Rather, the typically addressed problem is the modelling of vessel traffic by means of fuel savings and flow prediction through both neural networks (Haiyan and Youzhen, 2015) and weighted graphs (Hang et al., 2015). Image processing techniques have also been used for marine traffic applications, although most aimed to detect anomalies (Kraiman et al., 2001; Riveiro, 2011) or vessel movements (Willems et al., 2009). Other works are related to high-risk spot identification from an AIS reports database (Chang et al., 2010) as well as to traffic flow composition and vessel arrival or passage time. Within this paper we do not need to take account of the time parameter since the main goal is the evaluation, at a given time, of the mean number of nearby vessels. Of course this number depends on temporal traffic flow evolution but, from a radar design point of view, only the maximum value is significant, i.e. the value related to the time instant with the maximum number of sailing vessels.

Knowing the mutual distance distribution will also lead to a kind of marine traffic topology classification, i.e. to decide whether ships, at a given time, are following a route or whether they are "randomly" sailing over a finite region. An interesting related work by Su and Chang (2008) aimed at detecting the presence of fishing vessel spatial clusters for over-fishing prevention. The added value of the mutual distance approach is the detection of routes in addition to the distributed clusters.

The main goal of this paper is to classify traffic topology by using a single statistical parameter estimated from the observed real-world AIS data. Figure 1 presents the workflow chart of the proposed statistical approach.
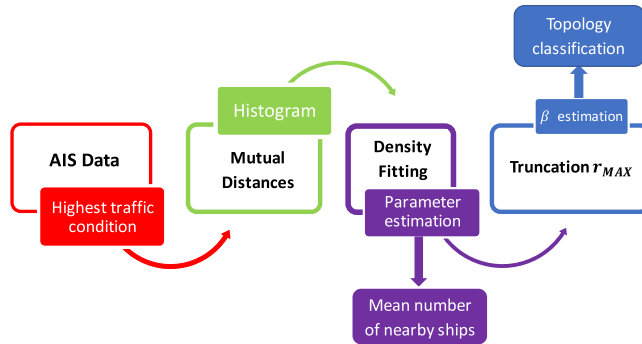
Figure 1. Workflow chart: from the AIS data to the traffic topology classification.

Starting from the full AIS reports database, the mean number of nearby ships is obtained, finally the traffic topology is classified through a single parameter (β) estimation.

In Section 2 AIS data is shown with the related statistical analysis in which the parameters of the Gamma and Generalised Gamma models are estimated. Section 3 considers the truncation of the statistical distribution of the mutual distances in order to evaluate the mean number of ships in a given region. The relationship between the statistical model parameters and the topology of the traffic has been investigated. In the Appendix a more general theoretical Poisson-like model is described for the sake of comparison. Note that all distances are in nautical miles (NM).

2. THE MARINE TRAFFIC MODEL. In this section the statistical model for the mutual distances is derived from AIS data.

2.1. *AIS data and their distribution.* The General Command of the Italian Coast Guard kindly provided the AIS data for the week 23 February − 1 March 2015 related to six areas: (1) Central Adriatic, (2) Otranto Canal, (3) Central Tyrrhenian, (4) Messina Strait, (5) Canal of Sicily and (6) Dardanelles/Bosphorus (see Figure 2 and Table 1(a) for more details).

Each area was originally sampled at a rate of one minute, then the huge AIS reports database has been reduced to a rate of four hours from midnight. We have taken the four-hour interval in which maximum traffic appears, neglecting what happens in each other four-hour time period. From data inspection, however, we did not perceive a significant deviation from the maximum traffic situation within the interval, so our analysis is conservative (Galati et al., 2015; Galati and Pavan, 2015).

From the first analysis of the AIS data, we derived the time interval having the maximum number of ships in each area, as shown in Table 1(b). In the following we refer to the area with the highest traffic as the area with the highest number of ships. The density z [ships/NM^2] of en-route ships is computed as the number of ships over the sea surface (Table 1(a)) in the highest traffic condition. We extrapolated the position of ships from the AIS data related to Table 1(b) (i.e. highest traffic condition) for each area. We used the flat earth approximation for distance due to the small-sized areas (max distance in area (6) is about 370 NM).
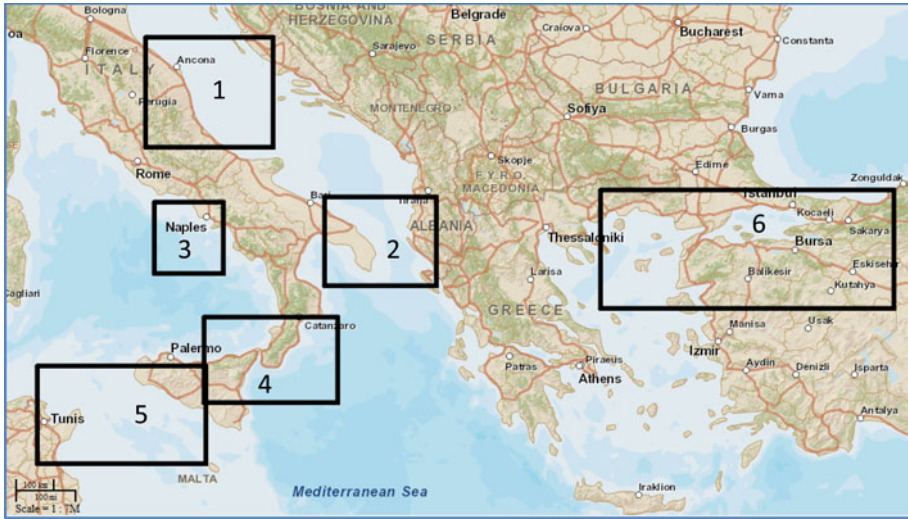
Figure 2.   View of the six Mediterranean areas.

Table 1(a).   Main characteristics of the six areas.

Area	N-E Point	S-O Point	Total Surface [NM^2]	Sea Surface [NM^2]	Sea [%]
(1)	44°10'18.40"N	42°09'26.58"N	22632	13600	60
Central Adriatic	15°55'16.71"E	12°43'13.25"E	(164 x 138)		
(2)	41°12'57.47"N	39°31'42.97"N	17712	12300	69
Otranto Canal	20°01'18.74"E	17°12'28.32"E	(144 x 123)		
(3)	41°07'27.98"N	39°46"07.02"N	8455	6700	79
Central Tyrrhenian	14°40'34.17"E	12°55'19.09"E	(95 x 89)		
(4)	38°55'08.47"N	37°13'27.60"N	20384	13700	67
Messina Strait	17°33'00.99"E	14°10'22.01"E	(182 x 112)		
(5)	37°56'26.98"N	35°59'03.12"N	30186	22800	75
Canal of Sicily	14°14'01.89"E	09°56'44.44"E	(258 x 117)		
(6)	41°21'26.79"N	39°05'16.24"N	60112	21700	36
Dardenelles Bosphorus	31°32'03.49"E	24°09'53.99"E	(442 x 136)		

Table 1(b).   Maximum number of ships per each area and their density z. Data for the week 23 February – 1 March 2015.

Area	Day and Time (in Feb 2015)	Max number of ships, N	Ships' density $z \left[\dfrac{ships}{NM^2} \times 10^{-3} \right]$
(1) Central Adriatic	Tue 24 04:00	285	20.88
(2) Otranto Canal	Tue 24 08:00	46	3.74
(3) Central Tyrrhenian	Fri 27 08:00	45	6.72
(4) Messina Strait	Fri 27 16:00	74	5.40
(5) Canal of Sicily	Fri 27 08:00	104	4.56
(6) Dardenelles/Bosphorus	Thu 26 12:00	53	2.44

Table 2(a).   Estimated parameters of the Gamma model for the six areas.

Area	Gamma Model		$\hat{m}_R = \dfrac{\hat{b}_{ML}}{\hat{\lambda}_{ML}}$ [NM]	$\dfrac{\hat{m}_R}{z}$
	$\hat{b}_{ML}$	$\hat{\lambda}_{ML}[\text{NM}^{-1}] \times 10^{-3}$		
(1)	2.1542	38.7	55.66	2.66
(2)	1.9371	47.2	41.04	11.0
(3)	2.0472	62.4	32.80	4.88
(4)	2.4059	42.9	56.08	10.4
(5)	1.8674	34.7	53.81	11.8
(6)	1.5753	27.4	57.50	23.5

Table 2(b).   Estimated parameters of the Generalised Gamma model for the six areas.

Area	Generalised Gamma Model			$\hat{m}_R = \dfrac{\Gamma(\hat{b}_{ML} + \frac{1}{\hat{\mu}_{ML}})}{\hat{\lambda}_{ML}\Gamma(\hat{b}_{ML})}$
	$\hat{b}_{ML}$	$\hat{\mu}_{ML}$	$\hat{\lambda}_{ML}[NM] \times 10^{-3}$	
(1)	0.6061	2.287	11.9	55.63
(2)	0.8303	1.709	19.1	41.01
(3)	0.3334	3.576	16.5	33.04
(4)	0.3939	3.525	10.6	55.89
(5)	0.3848	3.03	10.3	53.63
(6)	0.7918	1.559	13.1	57.63

The number of mutual distances is:

$$N = \frac{n \cdot (n-1)}{2} \tag{1}$$

in which n is the total number of ships in the area in a given time interval (e.g. with the highest traffic condition). It is worth noting that the N distances are not statistically independent because they are among ships: given n ships, if only one of them is moved, $n-1$ distances do change.

2.2.   *Statistical analysis of inter-ship distances.*   The ship-to-ship distance R can be fitted with a probability density function $f_R(r)$ having the following properties: $f_R(r) = 0$ for $r \leq 0$ and $\lim_{r \to \infty} f_R(r) = 0$.

A suitable candidate for this positive random variable is the Gamma model (Papoulis, 1990) defined by two parameters (i.e. the scale parameter λ and the shape parameter b):

$$f_R(r) = \frac{\lambda^b}{\Gamma(b)} r^{b-1} e^{-\lambda r} \ r \geq 0 \tag{2}$$

where $\Gamma(b) = \int_0^{+\infty} y^{b-1} e^{-y} dy$ is the Gamma function. To improve the model, a third parameter μ can be added into Equation (2) obtaining a Generalised Gamma model (Stacy, 1962):

$$f_R^{GEN}(r) = \frac{\mu \cdot \lambda^{b\mu}}{\Gamma(b)} r^{b\mu-1} e^{-(\lambda r)^{\mu}} \ r \geq 0 \tag{3}$$

When $\mu = 1$ the Generalised Gamma density function equals the Gamma model. These parameters can be estimated by the Maximum Likelihood (ML) method, which leads to a
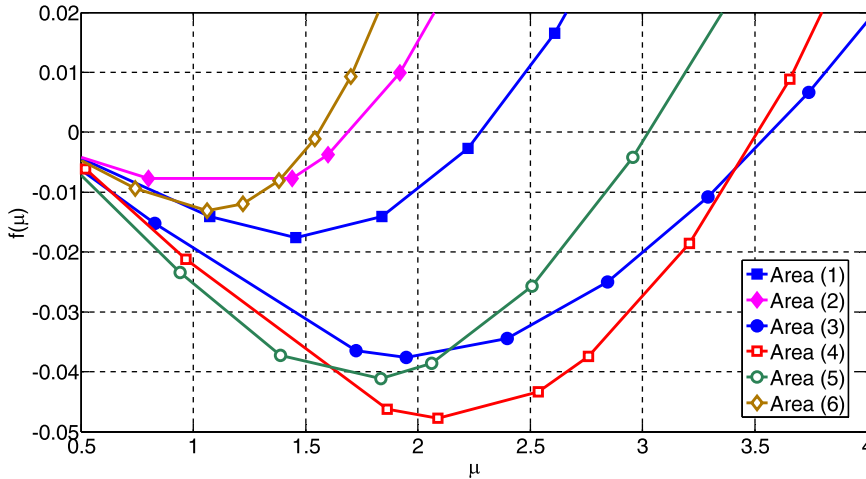
Figure 3.   The derivative of the Likelihood function for the estimation of the Generalised Gamma parameter μ. The $\hat{\mu}_{ML}$ is obtained posing $f(\mu) = 0$.

system of non-linear equations whose solutions are the values in Table 2(a) and Table 2(b), where the estimated mean values $\hat{m}_R$ (in NM) are shown.

The sample size for each area varies from 990 distances with average value of 32.80 NM (Area 3) to 40,470 distances with average value of 55.66 NM (Area 1). Day and time are listed in Table 1(b). For the Gamma model the ratio $\frac{\hat{m}_R}{z}[\frac{NM^3}{ships}]$ gives an idea about the topology of the traffic on the considered sea surface (e.g. en-route or randomly distributed): low ratio values correspond to a distributed, or random, topology (i.e. Central Adriatic, Area (1)), while higher values are related to a route or more regular topology (for example, in Otranto Canal (2), Messina Strait (4) and Canal of Sicily (5)). Moreover, we observe that the ML estimation of μ leads to a system of three non-linear equations where the μ-th power of the sample values (i.e. the measured distances) is present. Therefore, it is necessary to find that value of μ whereby the derivative of the Likelihood function, $f(\mu)$, is equal to zero (see Figure 3). However, as shown in Figure 3, the values of $f(\mu)$ in the range of practical interest, i.e. $0 < \mu < 3$, are close to zero, i.e. there are sub-optimal solutions (values of $\hat{\mu}$) that can be considered, including $\hat{\mu} = 1$. The use of $\hat{\mu} = 1$ simplifies the model leading back to the Gamma model that looks more convenient than its generalisation (see also in the following).

To validate the estimated parameters $\hat{b}_{ML}, \hat{\mu}_{ML}, \hat{\lambda}_{ML}$ the Kolmogorov-Smirnov test and the χ^2 test (Papoulis, 1990) should be applied with the null hypothesis being (respectively for the Gamma and the Generalised Gamma distribution): $H_0 : F(r) = F_R(r)$ or $H_0 : F(r) = F_R^{GEN}(r)$. However, since the N distances are not statistically independent, the tests too often reject the null hypothesis H_0 (Gleser and Moore, 1983), and cannot be effectively applied in the present case. However, a visual inspection gives a fairly good idea of the goodness of fit of the measured mutual distances with these distributions.

In Figures 4(a)–4(f) the histograms of distances for all areas are presented with the overlapped Gamma and Generalised Gamma estimated models. In some cases, e.g. Areas (3) and (5), the Generalised Gamma model is not the best fit because the third parameter μ improves the fit only locally. Hence, a first result of this study is that when considering
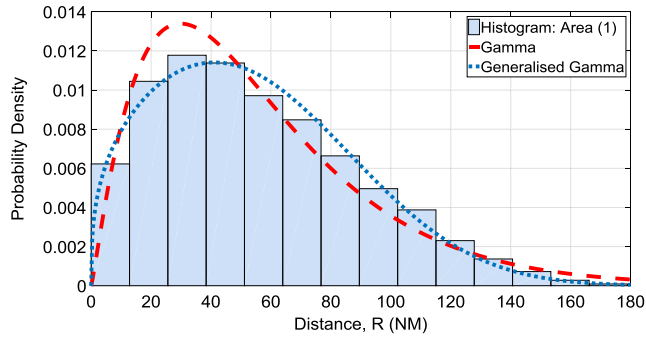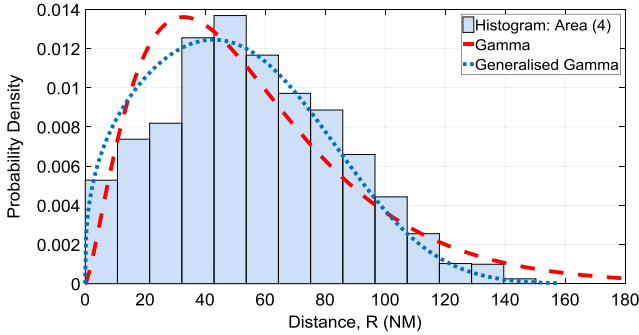
Figure 4(a).   Histogram and densities of R for Area (1).
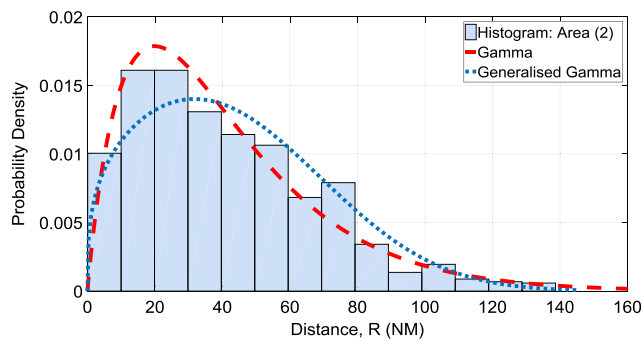
Figure 4(b).   Histogram and densities of R for Area (2).
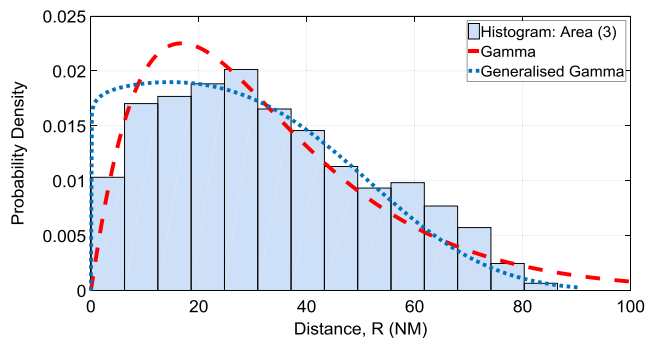
Figure 4(c).   Histogram and densities of R for Area (3).

the ship-to-ship distances a three parameters distribution such as the Generalised Gamma does not practically improve the goodness of fit with respect to a simpler, straightforward distribution such as the Gamma. Therefore, the Gamma model with parameters λ and b will be used in this work.

Figure 4(d).   Histogram and densities of R for Area (4).
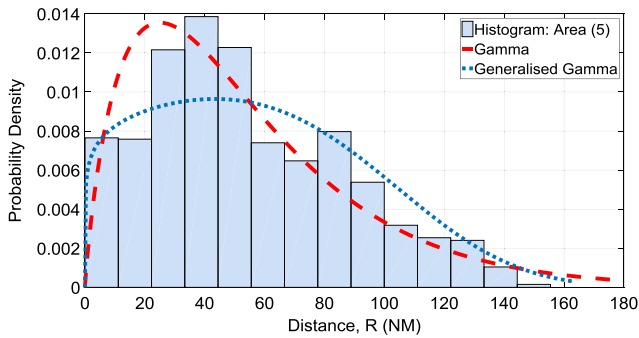

Figure 4(e).   Histogram and densities of R for Area (5).
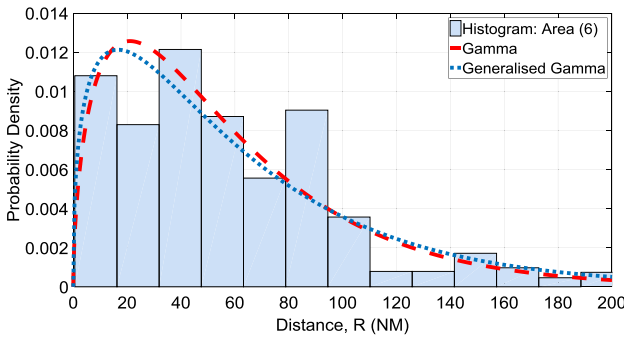

Figure 4(f).   Histogram and densities of R for Area (6).

3. VISIBILITY.   In the previous section we have shown that the distances between pairs of ships can be represented by a random variable R distributed according to a Gamma model with parameters b and λ.

As explained before, it is useful to consider, for a generic ship, the mean number of vessels in its surroundings within a specific area (Galati et al., 2015). The radar horizon – with the 4/3 earth propagation model – and the heights of ships must be considered to compute the maximum distance at which two on board radars may interfere. Radar horizon
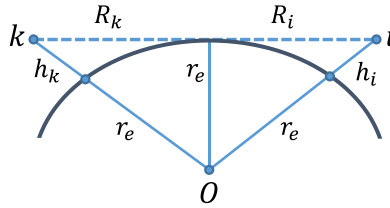
Figure 5. Radar horizon between ships k and i.

for two ships labelled "i" and "k" is related to the heights of on board radars h_k and h_i (see Figure 5).

In the standard atmosphere, making use of the equivalent earth radius $r_e = \frac{4}{3} r_{earth} \cong$ 8500 km, the horizon R_{ki} is:

$$R_{ki} = R_k + R_i \cong \sqrt{2 r_e} \cdot (\sqrt{h_k} + \sqrt{h_i}) \tag{4}$$

The antenna height, not being included in AIS data, must be estimated. Based on the AIS information of the ship and on photos of the various types of ship, we have empirically estimated the relationship between the length (as provided by AIS) of the ship and the radar antennae height (Galati et al., 2015). For example, if we consider two cargo ships with their radar antennae at 30 m above sea level, the optical horizon is about r_{MAX} = 35 NM, while it becomes r_{MAX} = 10 NM for small and pleasure boats, because of the antenna heights of the order of 4 m. In this section we focus only on the latter case (r_{MAX} = 10 NM). Let us consider an all-sea circular section with diameter r_{MAX}. It is possible to calculate the average number of ships (randomly) distributed in this circular sea surface through the probability P that the mutual distances do not exceed r_{MAX}:

$$P\{R \le r_{MAX}\} = \frac{1}{\Gamma(b)} \int_0^x e^{-t} t^{b-1} dt = \gamma(b, x) \tag{5}$$

where $\gamma(b, x)$ is the Incomplete Gamma Function (Abramowitz and Stegun, 1964) with $x = \lambda \cdot r_{MAX}$. The parameters b and λ have been estimated with the Maximum Likelihood method for each area (Tables 2(a) and 2(b)). Multiplying the probability in Equation (5) by the total number of ships in the area (n_{TOT}) we obtain the expected number of ships inside this circular sea surface:

$$n_{ships} = P\{R \le r_{MAX}\} \cdot n_{TOT} \tag{6}$$

The probability density of the random variable R, i.e. the mutual distances among the n_{ships} vessels in the area (with $0 \le R \le r_{MAX}$) is given by the conditional density function of Equation (2):

$$f_R(r | R \le r_{MAX}) = \begin{cases} \dfrac{f_R(r)}{F_R(r_{MAX})} & 0 < r < r_{MAX} \\ 0 & r \ge r_{MAX} \end{cases} \tag{7}$$

This approach would use the same parameters as estimated for the original model and therefore might not be fully reliable. Using Equation (7) to compute the conditional density

Table 3.   Estimation of b, μ, λ for $r_{MAX} = 10\,\mathrm{NM}$.

Area	Truncated Gamma		Truncated Generalised Gamma		
	$\hat{b}_{ML}$	$\hat{\lambda}_{ML}[\mathrm{NM}^{-1}]$	$\hat{b}_{ML}$	$\hat{\mu}_{ML}$	$\hat{\lambda}_{ML}[\mathrm{NM}^{-1}]$
(1)	1.46	9.3×10^{-12}	1.46	1	9.5×10^{-14}
(2)	1.58	2.7×10^{-12}	1.59	1	2.7×10^{-12}
(3)	1.25	3.6×10^{-12}	1.26	1	3.7×10^{-12}
(4)	1.02	0.012	0.19	5.25	6.9×10^{-4}
(5)	0.99	0.017	1.21	0.82	0.015
(6)	1.74	0.078	0.22	7.10	0.09

model from the Gamma model with parameters b, λ we can readily obtain:

$$f_R(r|R \leq r_{MAX}) = \begin{cases} \dfrac{\lambda^b x^{b-1} e^{-\lambda x}}{\gamma(b, \lambda r_{MAX})} & 0 < r < r_{MAX} \\ 0 & r \geq r_{MAX} \end{cases} \qquad (8)$$

In Equation (8) we have added the third parameter r_{MAX} named "truncation parameter" which takes into account the maximum distance at which the model should be considered (e.g. the optical horizon).

To estimate b and λ in Equation (8), having fixed the value of r_{MAX}, a closed-form solution such as the well-known one for the Gamma and Generalised Gamma distributions does not exist. The problem of finding the maximum for the Likelihood function must be solved by a non-linear optimisation method. We have used the Nelder-Mead algorithm (Nelder and Mead, 1965). This estimation often gives very low values for λ, as shown in Table 3 for Areas $(1) - (4)$.

For this reason a different model, when $\lambda \to 0$, has been considered for the short range distance between a pair of vessels, i.e. $r < r_{MAX}$, having set $r_{MAX} = 10\,\mathrm{NM}$.

3.1.   *Short Range Model.*   In Equation (8) when λ is close to zero, the only significant remaining term is x^{b-1} multiplied by a constant c depending on b. Posing $\beta = b - 1$ we obtain:

$$f_R(r|R \leq r_{MAX}) = c \cdot x^{\beta} \quad 0 < r < r_{MAX} \qquad (9)$$

The unity area condition for Equation (9) leads to:

$$c = \frac{\beta + 1}{r_{MAX}^{\beta+1}} \qquad (10)$$

Therefore, the conditional density for truncated distances with a single parameter β is:

$$f_R(r|R \leq r_{MAX}) = \begin{cases} \dfrac{\beta + 1}{r_{MAX}^{\beta+1}} \cdot r^{\beta} & 0 < r < r_{MAX} \\ 0 & r \geq r_{MAX} \end{cases} \qquad (11)$$

Normalising with respect to r_{MAX}, i.e. posing $\hat{r} = \frac{r}{r_{MAX}}$ in Equation (11) and multiplying Equation (11) for r_{MAX}, results in:

$$r_{MAX} \cdot f_R(r|R \leq r_{MAX}) = \begin{cases} (\beta + 1) \cdot \hat{r}^{\beta} & 0 < \hat{r} < 1 \\ 0 & r \geq 1 \end{cases} \qquad (12)$$
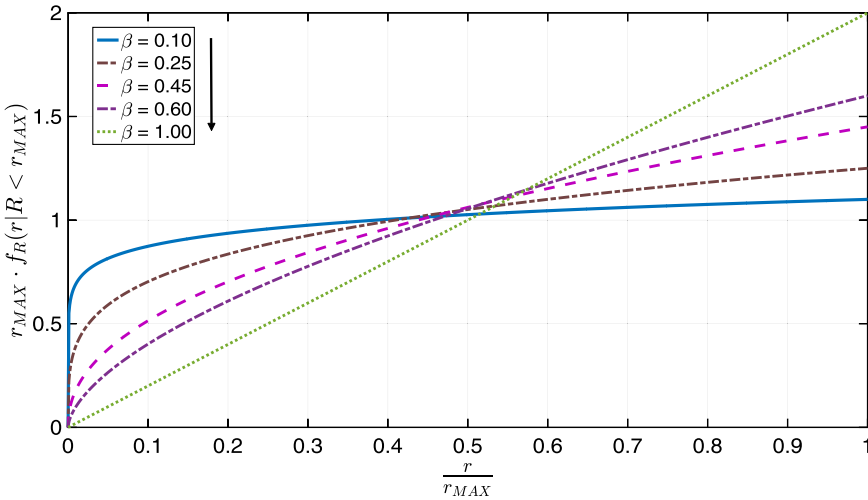
Figure 6.   Normalised conditioned density model.

Figure 6 shows Equation (12) for different values of β ($0 < \beta \leq 1$). If $\lambda \to 0$ (see Table 3) the Gamma model leads to Equation (11) and if $\beta \to 0$ the model converges to the uniform model:

$$f_R(r|R \leq r_{MAX}) = \frac{1}{r_{MAX}} \quad 0 < r < r_{MAX} \tag{13}$$

In the following, we show the results related to Central Adriatic, Canal of Sicily and Strait of Messina. It is known that the traffic in the Central Adriatic is mainly due to fishing boats whose positions include a large number of randomly distributed clusters. In the Canal of Sicily the vessel traffic can be classified as traffic near the coast (where normally fishing boats are the majority with some possible localised clusters) and traffic along the sea routes where cargo ships, container ships and tankers prevail. The difference in ship type leads us to evaluate the r_{MAX} limit: the bigger the ship, very likely the higher is its radar antenna, and so the wider is its interference area. Figure 7(a) and (b) show the estimated values of β using all data during the week 23 February – 1 March 2015 (42 values in total) versus the ship's density (number of ships per unit area) for Area (1) and Area (5) respectively.

From Figures 7(a) and 7(b) we can deduce the following:

(i) If the density of ships is much lower than 2.5×10^{-3} ships/NM^2 (in Central Adriatic it occurs mostly during the weekend when the fishery is strongly reduced), the low number of samples does not allow us to estimate β correctly.

(ii) For Area (1) increasing the density, i.e. when the density is greater than 2.5×10^{-3} ships/NM^2, the estimated β increases to around 0.5. For the maximum density of 20.88×10^{-3} ships/NM^2, β is 0.461 (dashed circle in Figure 7(a)); this case corresponds to the traffic shown in Figure 8 where the scenario is shown on the left (251 fishing boats and 33 other vessels). On the right the estimated truncated density function is plotted. The latter agrees with the data (see histogram).
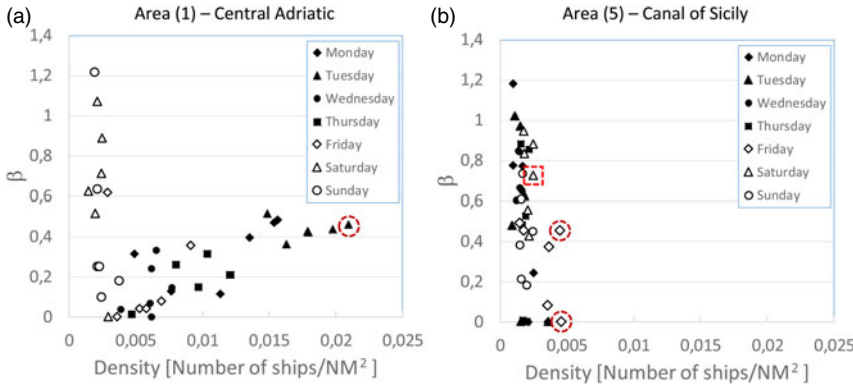
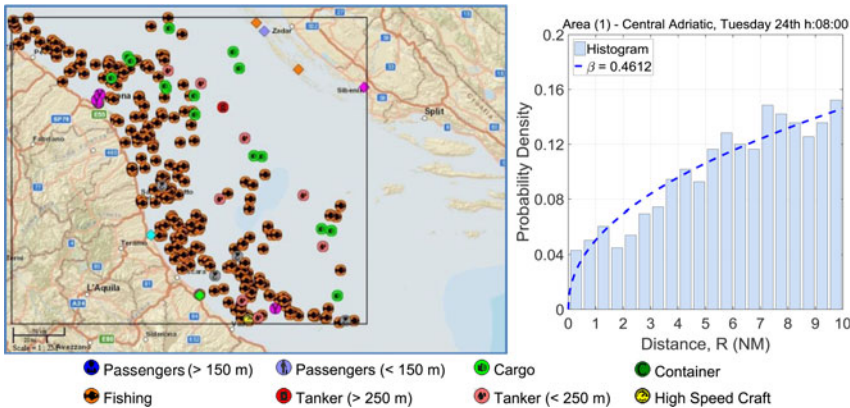Figure 7.   Estimated values of : (a) Area (1); (b) Area (5).

Figure 8.   Scenario of traffic for Area (1) Central Adriatic, Tuesday 24 at 08:00 and the estimated truncated density function. Fishing boats: 251 (88 %), others: 33 (12 %).

(iii) For Area (5) the density always remains below 5×10^{-3} ships/NM^2. Considering the two higher observed densities, i.e. 4.56×10^{-3} and 4.47×10^{-3} ships/NM^2, corresponding to the traffic shown in Figures 9(a) and 9(b) respectively, the estimated β is 6.1×10^{-8} and 0.4558 (dashed circles in Figure 7(b)). The different conditioned distribution is due to the fact that in the first case (h: 08:00), in comparison with the second one (h: 00:00), the presence of a greater number of fishing vessels (close to each other) implies an increase of the short ranges into the interval (0, $r_{MAX} = 10$ NM). In the second case the decrease of fishing boats in a few localised clusters and the increase of the others (cargo ships, container ships and tanks) cause an increase of the higher ranges as shown in Figure 9(b).

For the Canal of Sicily, a third case (Saturday 28 at 08:00) is shown in Figure 9(c) when cargo ships, container ships and tankers are predominant in comparison with the fishing boats. The density of ships is low, $z = 2.46 \times 10^{-3}$ ships/NM^2, and no clusters of fishing vessels are present. The other vessels, i.e. cargo ships, container ships and tankers, are more uniformly distributed along the routes. The resulting value of $\beta = 0.7265$.
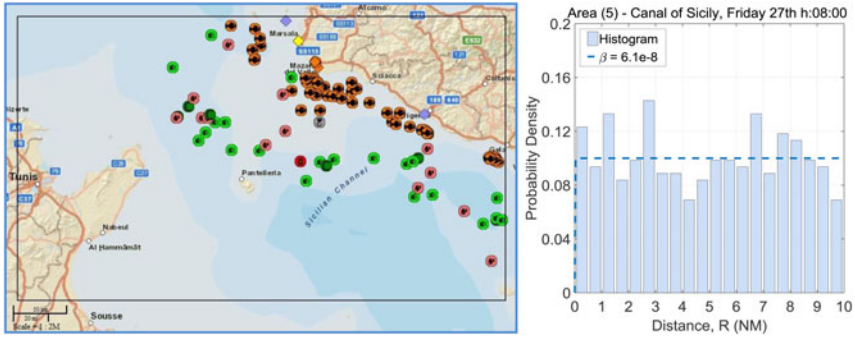
Figure 9(a).   Distributed traffic of Area (5) Canal of Sicily, Friday 27 at 08:00 and estimated truncated density function. Fishing boats: 62 (60%), others: 42 (40%).
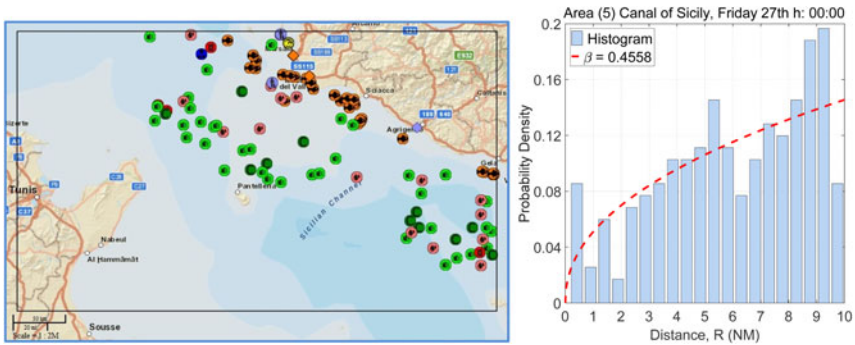

Figure 9(b).   Distributed traffic of Area (5) Canal of Sicily, Friday 27 at 00:00 and estimated truncated density function. Fishing boats: 31 (30%), others: 71 (70%).
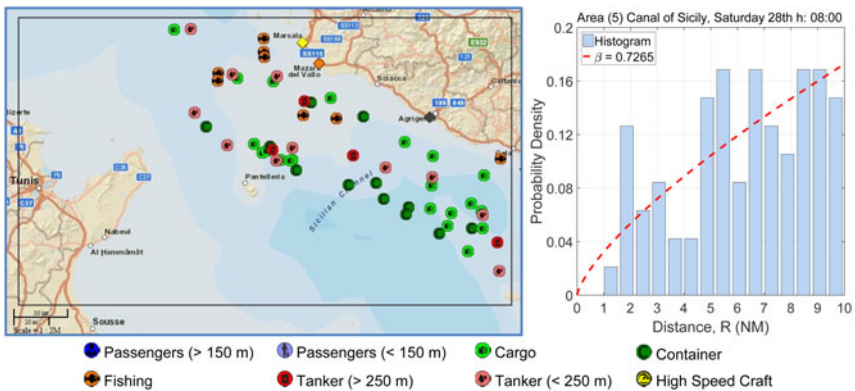

Figure 9(c).   Distributed traffic of Area (5) Canal of Sicily, Saturday 28 at 08:00 and estimated truncated density function. Fishing boats: 8 (14%), others: 48 (86%).
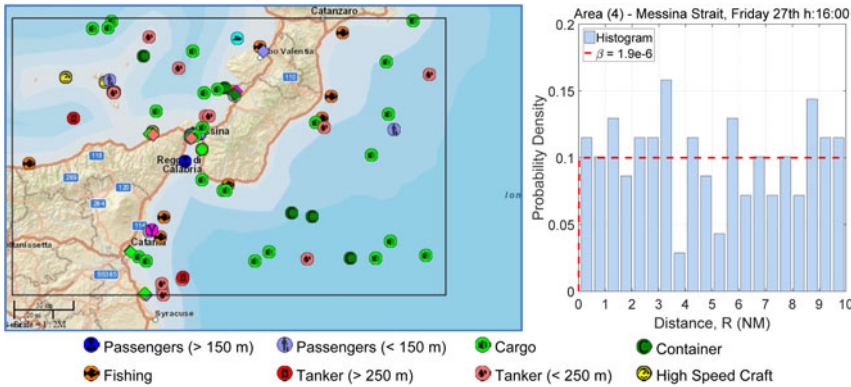
Figure 10.   Distributed traffic of Area (4) Messina Strait, Friday 27 at 16:00 and estimated truncated density function. Fishing boats: 9 (12%), others: 65 (88%).

Table 4.   Maximum Likelihood Estimate of β with $r_{MAX} = 10$ NM.

	Area (1)	Area (2)	Area (3)	Area (4)	Area (5)	Area (6)
$\hat{\beta}_{ML}$	0.461	0.589	0.257	1.9×10^{-6}	6.1×10^{-8}	0.455

The traffic in the Strait of Messina (Friday 27 at 16:00) is shown in Figure 10, where it is mainly due to large vessels and no localized clusters of fishing boats are present. The parameter β resulting is close to zero.

Table 4 reports the ML estimate of β for the six marine areas considering the maximum observed density of ships for each area with $r_{MAX} = 10$ NM.

From Table 4 we can find very low values for β in Areas (4) and (5). It is worth noting that in Area (6) β is comparable with that in the Central Adriatic although the area provides a main route. This effect is due to the presence, in Area (6), of two different seas (Aegean Sea and Sea of Marmara) as well as of the Dardanelles, one of the world's narrowest straits used for international navigation, with the likely effect of strongly distorting the behaviour of ships' distances with respect to the open sea. In general, the sea percentage in Table 1 also gives an idea about the reliability of the β values.

From the previous results we can maintain that when the density of ships is high enough to validate the results, the values of β are strongly dependent on the topology of the observed traffic and on the geographic area: β close to 0.5 represents many clusters of ships uniformly distributed over the sea, mainly due to small boats near to the coast (fishing boats). Values of β close to zero denote the absence of groupings of small boats, i.e. in this case large vessels are present away from the coast along the routes. Special evaluations should be done near harbours or close to the straits as they occur near Messina or in the Dardanelles areas.

In the Appendix, for comparison, a mathematical model is shown, based on a general bi-dimensional Poisson model. However, the use of the model is limited by the conditions of the real scenario that imposes a different topology of traffic (near to and far from the coast) that are non-uniformly distributed over the sea and, sometimes, strong limitations due to the geography of the area (presence of coasts, islands, straits, canals and harbours).

4. CONCLUSIONS. A full AIS report database has been provided by the Italian Coast Guard for six marine areas of the Mediterranean Sea. The sampled real-world AIS data have been used to create a statistical model for vessel-to-vessel mutual distances. A Gamma distribution fits the empirical vessel-to-vessel mutual distances well while the more complicated Generalised Gamma distribution does not improve the goodness of fit.

The need to know the mutual distance distribution arises due to the advent of the new solid state pulse compression marine radars whose operation is based on long pulses that are very likely to time-overlap each other. One promising radar interference rejection technique exploits the matched filter's pulse compression and provides for the transmission of several orthogonal radar signals, one for each radar. The number of needed orthogonal signals must be evaluated according to the number of nearby interfering ships. The statistical distribution of the mutual distances allows us to find it provided the area is limited to a maximum value r_{MAX}, which in turn is related to the ships' type size which is strongly correlated to the radar antenna height. The higher the radar antenna, the wider the interference area r_{MAX}, the higher the number of interfering vessels. In this paper we have mostly focused on the small to medium vessels with $r_{MAX} = 10 \, \mathrm{NM}$.

By truncating the mutual distances distribution to r_{MAX} it is possible to get a type of traffic topology classification which is easily handled by the single parameter β. The said classification is aimed to qualitatively evaluate the way in which the ships are positioned at a given time, whether randomly distributed (e.g. $\beta \approx 0$) or in a line (e.g. $\beta \to 1$) following some route.

From a more theoretical point of view, a mathematical approach based on the 2D Poisson points model leads to a more complex, polynomial truncated density function which, for a truncation distance much less than the dimension of the marine area, converges to a triangular density function, i.e. the same as the empirical model with $\beta = 1$.

More generally, the model is not aimed to predict any marine traffic behaviour, nor to optimise routes and fuel consumption.

The presented work is preliminary. The found model might lead to a powerful analysis of marine radar interferences from a statistical point of view, also leading to a change in solid state radar design starting from the proposed use of the orthogonal signals as a possible interference rejection technique. Moreover, other significant case studies may be analysed to prevent the model being affected by a particular coast's morphology or fishing vessel behaviour.

REFERENCES

Abramowitz, M. and Stegun, I.A. (1964). *Handbook of mathematical functions*. New York Dover Publications.

Amato, F., Fiorini, M., Gallone, S. and Golino, G. (2010). Fully solid state radar for vessel traffic services. *11th International Radar Symposium (IRS), 2010*.

Chang, S.J., Hsu, G.Y., Yang, J.A., Chen, K.N., Chiu, Y.F. and Chang F.T. (2010). Vessel Traffic Analysis for Maritime Intelligent Transportation System. *2010 IEEE 71st Vehicular Technology Conference (VTC 2010-Spring)*, 16–19 May, doi: 10.1109/VETECS.2010.5493942

Galati, G. and Pavan, G. (2015). Mutual interference problems related to the evolution of marine radars. *Proceedings of the Intelligent Transportation Systems (ITSC), Las Palmas, Gran Canaria (Spain)*, 1785–1790.

Galati, G., Pavan, G. and De Palo, F. (2015). Interference problems expected when solid-state marine radars will come into widespread use. *Proceedings of the Eurocon 2015, Salamanca (Spain)*, 1–6.

Galati, G., Pavan, G. and De Palo, F. (2016). Compatibility problems related with pulse compression, solid-state marine radars. *IET Journal Radar, Sonar & Navigation*, 1–7, doi:10.1049/iet-rsn.2015.0400.

Gleser, L.J. and Moore, D.S. (1983). The effect of dependence on chi-squared and empiric distribution tests of fit. *The Annals of Statistics*, **11(4)**, 1100–1108.

Haiyan W. and Youzhen W. (2015). Vessel traffic flow forecasting with the combined model based on support vector machine. *2015 International Conference on Transportation Information and Safety (ICTIS)*, 25–28 June 2015, doi: 10.1109/ICTIS.2015.7232151

Hang W., Xu M., Chen X. and Zhou S. (2015) Vessel traffic flow prediction model based on complex network. *2015 International Conference on Transportation Information and Safety (ICTIS)*, 25–28 June 2015, doi: 10.1109/ICTIS.2015.7232104

IMO (1997). Resolution A.857(20). Guidelines for Vessel Traffic Services.

Kraiman, J.B., Arouh, S.L. and Webb, M.L. (2001). Automated anomaly detection processor. *In Proceedings of SPIE: Enabling Technologies for Simulation Science VI*, Orlando, FL, USA, 1 April 2001; Sisti, A.F., Trevisani, D.A., Eds.; pp. 128–137.

Nelander, A. and Tòth-Pàl, Z. (2009). Modular System Design for a New S- Band Marine Radar. *International Radar Conference – Surveillance for a Safer World, 12–16 October 2009, Bordeaux, France*.

Nelder, J.A. and Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, **7(4)**, 308–313.

Papoulis, A. (1990). *Probability and statistics*. Prentice-Hall.

Philip, J. (1991). The probability distribution of the distance between two random points in a box. MSC: Primary 60D05, Secondary 52A22. Department of Mathematics, Royal Institute of Technology, Stockholm, Sweden.

Riveiro, M. (2011). Visual Analytics for Maritime Anomaly Detection, *Ph.D. Thesis*, University of Örebro, Sweden. http://oru.diva-portal.org/smash/get/diva2:381336/FULLTEXT06.pdf.

Stacy, E.W. (1962). A generalization of the gamma distribution. *The Annals of Mathematical Statistics*, September **33**, 1187–1192.

Su, Y.Y. and Chang, S.J. (2008). Spatial Cluster Detection for the Fishing Vessel Monitoring Systems. *MTS/IEEE Kobe Techno-Ocean OCEANS 2008*, 8–11 April, doi: 10.1109/OCEANSKOBE.2008.4531048

Willems, N., Wetering, H.V.D. and Wijk, J.J.V. (2009). Visualization of vessel movements. *Computer Graphics*. Forum 2009, **28**, 959–966.

Zhang, B., Dong, H. and Liu, T. (2013). Research on Marine Solid State Radar and its Application. *2013 Third World Congress on Information and Communication Technologies (WICT). IEEE*, doi: 10.1109/WICT.2013.7113098.

APPENDIX. GENERAL POISSON MODEL

A possible mathematical model for the distribution of ships in marine traffic can start from the bi-dimensional points of Poisson inside a rectangle. It supposes that the ships are uniformly placed over the sea within a rectangle. Let us consider a rectangle with sides L_1 and L_2 ($L_2 < L_1$) and a pair of random points (P_1, P_2) inside it (see Figure A1). It is supposed that the coordinates (U_1, U_2) of each point P_i are uniformly distributed in $(0, L_1)$ and $(0, L_2)$ respectively.

Then the distance $R = \overline{P_1 P_2}$ has a probability density function given by (Philip, 1991):

$$f_R(r) = \frac{2\pi r}{L_1 L_2} - \frac{4r^2}{L_1 L_2^2} - \frac{4r^2}{L_1^2 L_2} + \frac{2r^3}{L_1^2 L_2^2} \quad 0 < r \leq L_2 \tag{A1}$$

$$f_R(r) = -\frac{2r}{L_1^2} - \frac{4r^2}{L_1 L_2^2} + \frac{4r}{L_1 L_2} arcsin\left(\frac{L_2}{r}\right) + \frac{4r}{L_1 L_2^2}\sqrt{r^2 - L_2^2} \quad L_2 < r \leq L_1 \tag{A2}$$

$$f_R(r) = -2r\left[\frac{1}{L_1^2} + \frac{1}{L_2^2}\right] + \frac{4r}{L_1 L_2}\left[arcsin\left(\frac{L_1}{r}\right) + arcsin\left(\frac{L_2}{r}\right)\right]$$
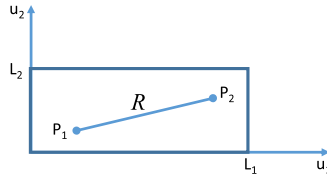
Figure A1. Two random points inside the rectangle ($L_2 < L_1$).

$$+ \frac{4r}{L_1 L_2} \left[\sqrt{\left(\frac{r}{L_1}\right)^2 - 1} + \sqrt{\left(\frac{r}{L_2}\right)^2 - 1} \right] - \frac{2\pi r}{L_1 L_2} - \frac{2r^3}{L_1^2 L_2^2}$$

$$L_1 < r \leq \sqrt{L_1^2 + L_2^2} \tag{A3}$$

$$f_R(r) = 0 \quad \text{elsewhere} \tag{A4}$$

Without loss of generality we pose $L_1 = cL_2$ with the constant $c \geq 1$.

For $c = 1$ and $c = 20$ the density functions $f_R(r)$ are shown in Figures A2(a) and A2(b) respectively (continuous line).

When $L_2 \ll L_1$ (at limit $L_2 \to 0$ with a finite value of L_1) the density $f_R(r)$ converges to a triangular shape (dashed line). In Figures A2(a) and A2(b) the histograms of data (obtained by simulation, i.e. randomly allocating a large number of vessels in the rectangular area (12,686 for Figure A2(a) and 12,294 for Figure A2(b)) and measuring their mutual distances) are also shown.

For $0 < r \leq L_2$ the density $f_R(r)$ can be rewritten as:

$$f_R(r) = \frac{2\pi r}{cL_2^2} - \frac{4r^2}{cL_2^3} - \frac{4r^2}{c^2 L_2^3} + \frac{2r^3}{c^2 L_2^4} \quad 0 < r \leq L_2 \tag{A5}$$

and the corresponding distribution function in the interval $(0, L_2)$ is:

$$F_R(r) = \frac{\pi r^2}{cL_2^2} - \frac{4r^3}{3cL_2^3} - \frac{4r^3}{3c^2 L_2^3} + \frac{r^4}{2c^2 L_2^4} \quad 0 < r \leq L_2 \tag{A6}$$

If we suppose $r_{MAX} \leq L_2$, the conditioned density to the event $\{R < r_{MAX}\}$ results:

$$f_R(r | R < r_{MAX}) = \frac{f_R(r)}{F_R(r_{MAX})} \quad 0 < r \leq r_{MAX} \tag{A7}$$

Substituting Equations (A5) and (A6) we obtain:

$$f_R(r | R < r_{MAX}) = \begin{cases} \dfrac{2\pi r - \frac{4r^2}{L_2}\left(1 + \frac{1}{c}\right) + \frac{2r^3}{cL_2^2}}{\pi r_{MAX}^2 - \frac{4r_{MAX}^3}{3L_2}\left(1 + \frac{1}{c}\right) + \frac{r_{MAX}^4}{2cL_2^2}} & 0 < r \leq r_{MAX} \\ 0 & \text{elsewhere} \end{cases} \tag{A8}$$

When $L_1 = L_2 = L$ ($c = 1$):

$$f_R(r | R < r_{MAX}) = \frac{2\pi r - \frac{8r^2}{L} + \frac{2r^3}{L^2}}{\pi r_{MAX}^2 - \frac{8r_{MAX}^3}{3L} + \frac{r_{MAX}^4}{2L^2}} \quad 0 < r \leq r_{MAX} \tag{A9}$$
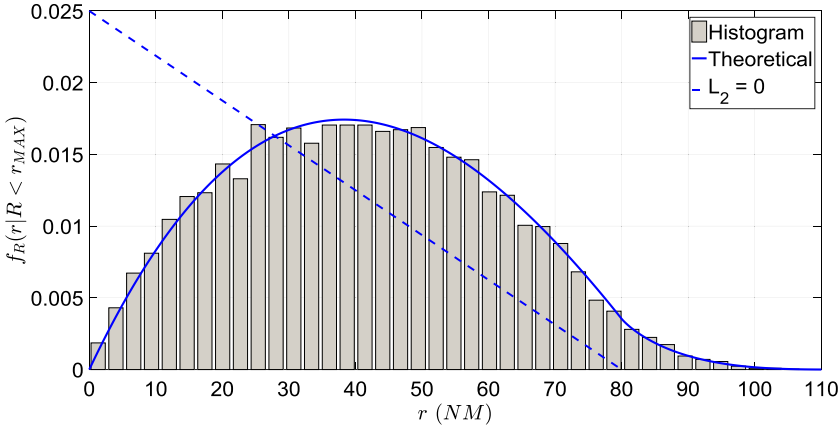
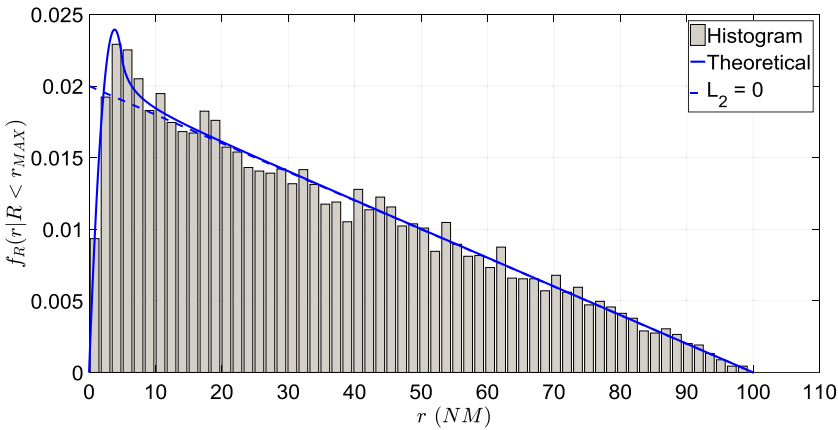Figure A2(a).   Density function of R for $c = 1$ ($L_1 = L_2 = 80$ NM).

Figure A2(b).   Density function of R for $c = 20$ ($L_1 = 100$ NM, $L_2 = 5$ NM).

Posing $L = \frac{r_{MAX}}{\alpha}$ with $0 < \alpha \leq 1$

$$f_R(r|R < r_{MAX}) = \frac{2\pi r - \frac{8\alpha}{r_{MAX}}r^2 + \frac{2\alpha^2}{r_{MAX}^2}r^3}{\pi r_{MAX}^2 - \frac{8}{3}\alpha r_{MAX}^2 + \frac{1}{2}\alpha^2 r_{MAX}^2} \quad 0 < r \leq r_{MAX} \tag{A10}$$

and normalising with respect to r_{MAX}, i.e. posing $\hat{r} = \frac{r}{r_{MAX}}$ and multiplying Equation (A10) by r_{MAX}:

$$r_{MAX} \cdot f_R(r|R < r_{MAX}) = \frac{2\pi \hat{r} - 8\alpha \hat{r}^2 + 2\alpha^2 \hat{r}^3}{\pi - \frac{8}{3}\alpha + \frac{1}{2}\alpha^2} \quad 0 < \hat{r} \leq 1 \tag{A11}$$

For $\alpha \to 0$, i.e. $r_{MAX} \ll L$, Equation (A11) converges to the straight line:

$$\lim_{a \to 0} r_{MAX} \cdot f_R(r|R < r_{MAX}) = 2\hat{r} \quad 0 < \hat{r} \leq 1 \tag{A12}$$

This means that when $r_{MAX} < L$ (i.e. for a relatively large sea traffic area, as compared to the visibility distance of pairs of vessels) the (truncated) density function is the same as Equation (12) with $\beta = 1$, see Figure 5, i.e. a truncated Gamma with $b = 2$.
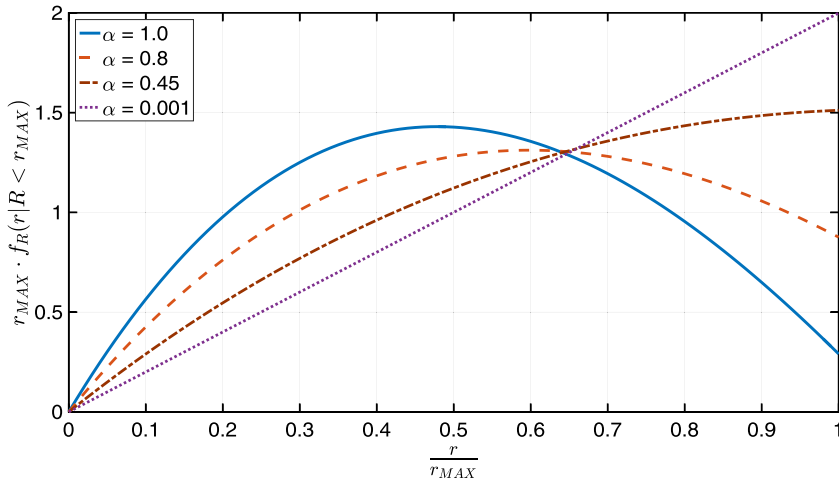
Figure A3.   Normalised conditioned density function with $L = \alpha \cdot r_{MAX}$.

While for $\alpha \to 1$ ($L = r_{MAX}$) Equation (A11) becomes:

$$\lim_{a \to 1} r_{MAX} \cdot f_R(r|R < r_{MAX}) = \frac{2\pi\hat{r} - 8\hat{r}^2 + 2\hat{r}^3}{\pi - \frac{13}{6}} \quad 0 < \hat{r} \le 1 \tag{A13}$$

with $r_{MAX} \cdot f_R(r|R < r_{MAX}) = 0$ elsewhere.

Figure A3 shows Equations (A11), (A12) and (A13) varying $\alpha = 1, 0.8, 0.45, 0.001$, and must be compared with Figure 5.