Dempster-Shafer Theory Based Ship-Ship Collision Probability Modelling

Alejandro Talavera Ortiz, Ricardo Aguasca Colomo, and Blas J. Galván-González

University Institute for Intelligent Systems
and Numerical Applications in Engineering
University of Las Palmas de Gran Canaria
Edificio Central del Parque Científico-Tecnológico
Campus Universitario de Tafira
35017 - Las Palmas de Gran Canaria
Spain
`atalavera@siani.es`

Abstract. The methodology proposed in this paper considers the uncertainty present in modeling the probability of collision between ships on a route. The proposal allows representing and quantifying uncertainty, and ensures rigorous propagation of this uncertainty from the input variables to the output variable.

This proposal complements the analysis of risk and helps the decision maker to know the degree of confidence associated with the results of the analysis.

Pedersen's model has been selected to estimate the probability of collision, using the information provided by the AIS, and Dempster-Shafer Theory has been selected for the treatment of uncertainty.

This methodology has been applied to maritime traffic in the Canary Islands and has been validated using the Kullback-Leibler divergence. The results are consistent with those obtained with the software IWRAP recommended by IALA.

Keywords: Maritime traffic safety, collision probability, Dempster-Shafer Theory, Data AIS.

1 AIS System

The use of AIS has provided a major improvement in the information available on the maritime traffic in an area and has changed the way of developing the modeling of the probability of collision between ships on a route.

The AIS system, mandatory installation in the terms established by the SOLAS convention, provides data on the position, velocity, type and length of vessels navigating in a route. These data allow defining the trajectories of ships (the position data of each vessel are sorted chronologically). It is also possible to model traffic, defining routes resulting from the clustering of the trajectories in areas of high traffic density. Subsequently with the use of the tables proposed by Lloyd's Register is possible to complete the description of maritime traffic. These tables set the index length/beam, average speed and average length in each segment length.

R. Moreno-Díaz et al. (Eds.): EUROCAST 2013, Part II, LNCS 8112, pp. 63–70, 2013.

2 Pedersen's Model

There are several reasons for selecting this model:

- Allows the grouping ships by type and size.
- Allows the use and analysis of AIS data.
- Incorporates lateral traffic distribution as an input variable.
- It is used by the computational model recommended by IALA, called IWRAP Mk2.

The model estimates the number of collision candidates (number of vessels exposed to a critical situation that may result in an accident) according to a surface integral extended to the crossing area and referred to all possible combinations of pairs of boats:

$$N_a = \sum_i \sum_j \iint_{\Omega(z_i z_j)} \frac{Q_{1i} Q_{2j}}{V_i^{(1)} V_j^{(2)}} f_i^{(1)}(z_i) f_j^{(2)}(z_j) V_{ij} D_{ij}\, dA \Delta t \qquad (1)$$

Where Q_{1i}, Q_{2j} represents the number of crossings for each class and length of ship in the direction (1) and (2), $V_i^{(1)}$is the speed of the ships of type i in the direction (1), $f_i^{(1)}$ is the lateral distribution of traffic of the ships of type i in the direction (1), and D_{ij} is geometrical collision diameter (meeting critical distance between two vessels).

In the case of parallel paths, this expression becomes:

$$N_a = L \sum_{i,j} P_{i,j} \frac{V_{ij}}{V_i^{(1)} V_j^{(2)}} \left(Q_i^{(1)} Q_j^{(2)}\right) \qquad (2)$$

Where L is the length of the path and $P_{i,j}$ represents the probability of collision between two vessels, which is obtained from the expression:

$$P_{i,j} = P\left[y_i^{(1)} + y_j^{(2)} < B_{ij}\right] - P\left[y_i^{(1)} + y_j^{(2)} < -B_{ij}\right] \qquad (3)$$

Where $y_i^{(1)}$ and $y_j^{(2)}$ represent the distances of the paths of the vessels to the axis of route and B_{ij} is the average length of the ships type i, j.

The probability of collision in the route is calculated by multiplying the number of candidates by a factor, called causal factor, P_c, which quantifies the possibility that the boats are not capable of making evasive maneuvers to avoid the accident. This causal factor is determined by the skills / abilities of the crew and the maneuverability of ships in accident situations. Therefore, this factor is independent of the traffic and is estimated by analysis of possible fault conditions or the study of historical data. Causal factor values generally used are those proposed by Fuji and Mizuki: $0{,}5 \cdot 10^{-4}$ for Head on collisions and $1{,}1 \cdot 10^{-4}$ for Overtaking collisions.

3 Uncertainty in the Pedersen's Model

The model inputs relating to a sample period are estimated with AIS information.

The uncertainty associated with these data is sample type, originated in the fact that we cannot say that the boats will behave in future periods the same way they have behaved in the sample period documented by the AIS.

This type of uncertainty leads to a context of incomplete numerical information, and may be represented by intervals. The Dempster-Shafer Theory, also called Theory of Evidence allows rigorous treatment of such uncertainty.

4 Dempster-Shafer Theory

This theory represents uncertain variables using Dempster-Shafer structures (DSS). DSS is a set of pairs formed by an interval and associated basic probability assignment.

The basic probability assignment associated with an interval expresses the amount of evidence supporting the claim that the variable of interest is contained in this interval. This basic probability assignment is defined on a universal set $\mathbb{X}$ as a function of the power set $\mathbb{P}_x$ in the interval [0,1].

From basic probability assignment are defined the two evidential measures: belief and plausibility. The measure of belief, Bel (X_i) for interval X_i represents the minimum belief in the statement that the variable of interest is contained in the interval X_i. The measure of plausibility, Pl (X_i) for interval X_i represents the maximum belief in the statement that the variable of interest is contained in the interval X_i.

Evidential measures can be calculated from the basic assignment of probability. For any intervals $X_i, X_j \in \mathbb{P}_x$:

$$Bel\,(X_i) = \sum_{X_j | X_j \subseteq X_i} m(X_j) \tag{4}$$

$$Pl\,(X_i) = \sum_{X_j | X_j \cap X_i \neq \emptyset} m(X_j) \tag{5}$$

The value of this theory lies in the probabilistic interpretation of evidential measures due to Dempster and Yager: "The boundaries of DSS make a probability box formed by the belief and plausibility functions. This probability box contains the distribution function of the variable under study”. This interpretation is completed by the duality between probability box and DSS and can move from one to another in the form proposed by Ferson.

5 Methodological Proposal

The DSS for the sample period is constructed from the histogram that shows the distribution of the crossings along the route. This DSS is formed by associating each interval with the relative frequency of crossings collected by the AIS system.

From this sample DSS, by application of the confidence limits of Kolmogorov-Smirnov, as proposed by Ferson, the probability box containing the set of future paths of the ships that sail on the route is obtained. To obtain these confidence limits, Ferson proposes to apply the expression:

$$Min\left(1, max\left(0, F(x) \pm D(\alpha, n)\right)\right) \tag{6}$$

Where $D(\alpha, n)$ is the Kolmogorov statistic, α is the confidence level and n is the number of intervals.

When the lateral distribution of future traffic in both directions of the route has been represented by a DSS, collision is considered possible (meeting situation) in the interval resulting from the intersection of these intervals which form DSS, (Fig. 1).

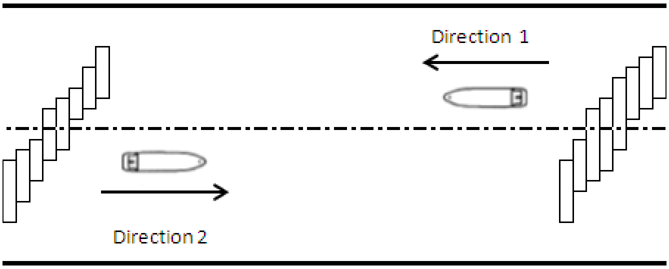


Fig. 1. Head on collision. Traffic represented with DSS.

With this new representation of the traffic, the probability of collision between two ships, $P_{i,j}$, is expressed as:

$$P_{i,j} = \left(B_i + B_j\right) \frac{\mathcal{L}\left[I_i^{(1)} \cap I_j^{(2)}\right]}{(b_i - a_i)(b_j - a_j)} \tag{7}$$

Where B_i is the beam of ship type i, $I_i^{(1)} = [a_i, b_i]$ is the interval i in the direction (1) , and $\mathcal{L}\left[I_i^{(1)} \cap I_j^{(2)}\right]$ is the length of the interval resulting from the intersection of the two intervals, calculated by interval arithmetic.

The equation that estimates the number of candidates for each pair of intervals can be written as:

$$N_a = L \sum_{i,j} P_{i,j} \frac{V_{ij}}{V_i^{(1)} V_j^{(2)}} Q_i^{(1)} h_i^{(1)} Q_j^{(2)} h_j^{(2)} \tag{8}$$

Where $h_i^{(1)}$ and $h_j^{(2)}$ are the frequencies of the voyages of ships and equal to the height of the different intervals that form the DSS.

The clustering and sorting of the resulting values, multiplied by the causal factor, allow obtaining the probability of collision in each of the intervals of the route according to the expression:

$$\lambda_{col} = N_a \cdot P_c \tag{9}$$

6 Application and Validation of the Proposal

The proposed methodology has been applied to maritime traffic in the Canary Islands in the period between August 2010 and August 2011 with data provided by the Ministry of Development.

By way of example, we present the results of applying the methodology proposed in this article and the methodology proposed by IWRAP to two routes: one that connects the Port of Santa Cruz de Tenerife and Puerto de Santa Cruz de La Palma, (TFE-SCP), and which connects the Port of Los Cristianos and the Port of La Gomera, (CRI-GOM), (Table 1, Table 2):

Table 1. Head on collision probability. Route TFE-SCP

Type of ships		Results proposed		Results IWRAP
Striking	Struck	Minimum	Maximum	
Oil products tanker	Oil products tanker	5.62741E-09	4.79340E-08	4.56873E-08
Oil products tanker	General cargo ship	3.12044E-08	2.65797E-07	2.53352E-07
Oil products tanker	Passenger ship	1.13202E-07	9.64245E-07	9.19210E-07
Oil products tanker	Other ship	1.79905E-10	1.53242E-09	1.27107E-09
General cargo ship	Oil products tanker	3.12044E-08	2.65797E-07	2.53352E-07
General cargo ship	General cargo ship	7.60681E-08	6.47944E-07	6.17463E-07
General cargo ship	Passenger ship	3.52714E-07	3.00439E-06	2.86699E-06
General cargo ship	Other ship	1.56583E-09	1.33377E-08	1.14392E-08
Passenger ship	Oil products tanker	1.13202E-07	9.64245E-07	9.19210E-07
Passenger ship	General cargo ship	3.52714E-07	3.00439E-06	2.86699E-06
Passenger ship	Passenger ship	1.47394E-06	1.25550E-05	1.19875E-05
Passenger ship	Other ship	5.16943E-09	4.40329E-08	3.71755E-08
Other ship	Oil products tanker	1.79905E-10	1.53242E-09	1.27107E-09
Other ship	General cargo ship	1.56583E-09	1.33377E-08	1.14392E-08
Other ship	Passenger ship	5.16943E-09	4.40329E-08	3.71755E-08
Sum		2.56371E-06	2.18375E-05	2.08295E-05

Table 2. Head on collision probability. Route CRI-GOM

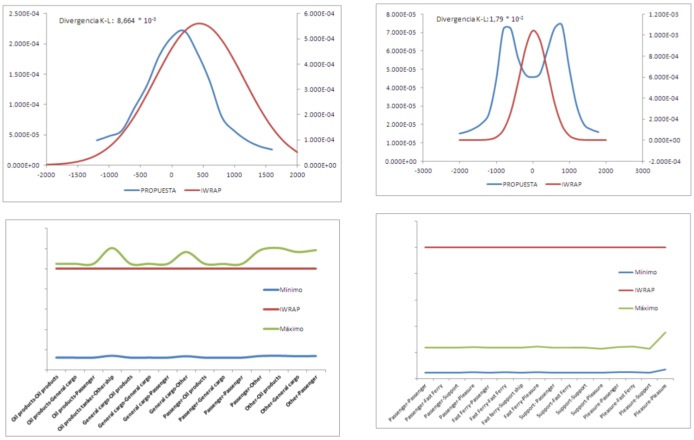
Type of ships		Results proposed		Results IWRAP
Striking	Struck	Minimum	Maximum	
Passenger ship	Passenger ship	1.53403E-05	7.62580E-05	3.21352E-04
Passenger ship	Fast Ferry	6.90269E-06	3.43266E-05	1.44392E-04
Passenger ship	Support ship	1.29122E-08	6.42113E-08	2.70445E-07
Passenger ship	Pleasure boat	2.68518E-08	1.33532E-07	5.49915E-07
Fast Ferry	Passenger ship	6.90269E-06	3.43266E-05	1.44392E-04
Fast Ferry	Fast Ferry	2.75458E-06	1.36984E-05	5.74276E-05
Fast Ferry	Support ship	5.89574E-09	2.93191E-08	1.23284E-07
Fast Ferry	Pleasure boat	1.18058E-08	5.87096E-08	2.38824E-07
Support ship	Passenger ship	1.29122E-08	6.42113E-08	2.70445E-07
Support ship	Fast Ferry	5.89574E-09	2.93191E-08	1.23284E-07
Support ship	Support ship	8.18784E-12	4.07176E-11	1.71476E-10
Support ship	Pleasure boat	1.86986E-11	9.29867E-11	4.08778E-10
Pleasure boat	Passenger ship	2.68518E-08	1.33532E-07	5.49915E-07
Pleasure boat	Fast Ferry	1.18058E-08	5.87096E-08	2.38824E-07
Pleasure boat	Support ship	1.86986E-11	9.29867E-11	4.08778E-10
Pleasure boat	Pleasure boat	3.23243E-11	1.60746E-10	4.53085E-10
Sum		3.20153E-05	1.59182E-04	6.69930E-04

The proposal made by IWRAP assumes that both traffic and collision probability follow a normal distribution, whereas the proposed methodology does not require the traffic follow any known distribution and it is represented by a DSS. The Kullback-Leibler divergence measures the discrepancy between two distribution functions and has been used for validation of the proposal. This metric is expressed as:

$$\sum_z P(z) \log_2 \frac{P(z)}{Q(z)} \tag{10}$$

The results obtained by the proposed methodology will approach those obtained by IWRAP insofar that the distribution function of the actual collision probability is set to the normal distribution proposed by IWRAP.

The graphical representation of the distribution functions obtained with both proposals and the Kullback-Leibler divergence associated are shown in the Fig. 2:



Route TFE-SCP Route CRI-GOM

Fig. 2. Validation of results

In the first case, Route TFE-SCP, IWRAP values are between maximum and minimum values obtained with the proposed methodology, with divergence values of the order of 10-3. By contrast, when the actual distribution does not fit a normal distribution, Route CRI-GOM, the divergence increases and the range of results obtained by the proposed methodology gets away from those proposed by IWRAP.

7 Conclusions and Applications

The proposed methodology takes into account the uncertainty present in the process and allows measuring the uncertainty associated with the results.

The proposal does not consider the model input variables follow some known distribution function, so it does not introduce possible mistakes if the actual function does not fit the expected.

Since the probability of collision is estimated for different sections of the route, security strategies oriented redirect traffic to less traffic areas may be established and also measure the improvement resulting from this strategy.

Some of the specific applications of the proposal are:

- The ability to study risk reduction if established safety corridors in the area.

- Studying the impact on the risk of collision of installing permanent structures in the area.

- Studying the effect on the risk of collision of the drift of a damaged ship in the area.

References

1. Dempster, A.P.: Upper and lower probabilities induced by a multi-valued mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
2. Yager, R.R.: Arithmetic and other operations on Dempster-Shafer structures. International Journal of Man-Machine Studies 25, 357–366 (1986)
3. Fuji, Y., Yamanouchi, H., Mizuki, N.: Some factors affecting the frequency of accidents in marine traffic. II: The probability of stranding. III: The effect of darkness on the probability of stranding. Journal of Navigation 27 (1974)
4. Mac Duff, T.: The probability of vessel collisions. Ocean Industry, 144–148 (1974)
5. Pedersen, P.T.: Collision and grounding mechanics. In: Proceedings of WEMT 1995. The Danish Society of Naval Architects and Marine Engineers, pp. 461–474 (1995)
6. Friis-Hansen, P., Engberg, P.C.: Basic modeling principles for prediction of collision and grounding frequencies. Technical University of Denmark (2008)
7. Montewka, J., Hinz, T., Kujala, P., Matusiak, J.: Probability Modelling of vessel collisions. Reliability Engineering and System Safety 95, 573–589 (2010)
8. Montewka, J., Goerlandt, F., Kujala, P.: Determination of collision criteria and causation factors appropriate to a model for estimating the probability of maritime accidents. Ocean Engineering 40, 50–61
9. Goerlandt, F., Kujala, P.: Traffic simulation based ship collision probability modeling. Reliability Engineering and System Safety 96(1), 91–107 (2011)
10. Fowler, T.G., Sorgard, E.: Modeling ship transportation risk. Risk Analysis 20(2), 225–244 (2000)
11. Sofartsstyrelsen. Risk analysis of sea traffic in the area around Bornholm. Technical report, COWI, Kongens Lyngby (2008)
12. Merrick, J., Van Dorp, J., Blackford, J., Shaw, G., Harrald, J., Mazzuchi, T.: A traffic density analysis of proposed ferry service expansion in San Francisco bay using a maritime simulation model. Reliability Engineering and System Safety 81, 119–132 (2003)
13. Merrick, J., Van Dorp, J.: Modeling risk in the dynamic environment of maritime transportation. In: Proceedings of the Winter Simulation Conference, pp. 1090–1098 (2001)
14. Gucma, L.: Models of ship's traffic flow for the safety of marine engineering structures evaluation. In: Proceedings of the ESREL, pp. 713–718 (2003)
15. Przywarty, M.: Probabilistic model of ships navigational safety assessment on large sea areas. In: Proceedings of the 16th International Symposium on Electronics in Transport (2008)
16. Shafer, G.: A mathematical theory of evidence. Princeton University Press (1976)
17. Sentz, K., Ferson, S.: Combination of evidence in Dempster-Shafer theory. Technical report. Sandia National Laboratories (2002)

18. Moore, R.E.: Interval Analysis. Prentice-Hall, Englewood Cliffs (1966)
19. Neumann, T.: A simulation environment for modeling and analysis of the distribution of shore observatory stations. International Journal on Marine Navigation and Safety of Sea Transportation 5(4) (2011)
20. Filipowicz, W.: An application of mathematical theory of evidence in navigation. International Journal on Marine Navigation and Safety of Sea Transportation 3(4) (2009)
21. Filipowicz, W.: Belief structures in position fixing. Communications in Computer and Information Science 104, 434–446 (2010)
22. Sii, H.S., Ruxton, T., Wang, J.: Synthesis using fuzzy set theory and a Dempster-Shafer based approach to compromise decision-making with multiple-attributes applied to risk control options selection. Proceedings Institution of Mechanical Engineers 216, 15–29 (2002)
23. Helton, J.C.: Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty. Journal of Statistical Computation and Simulation 57, 3–76 (1997)
24. Ferson, S., Kreinovich, V., Ginzburg, L., Myers, D., Sentz, K.: Constructing probability boxes and Dempster-Shafer structures. Technical report. Sandia National Laboratories (2003)
25. Miller, L.H.: Table of percentage points of Kolmogorov statistics. Journal of the American Statistical Association 51, 111–121 (1956)
26. Ferson, S., Hajagos, J.: Arithmetic with uncertain numbers: rigorous and (often) best possible answers. Reliability Engineering and System Safety 85, 135–152 (2004)
27. Ferson, S., Tucker, W.T.: Sensitivity in Risk analyses with uncertain numbers. Technical report. Sandia National Laboratories (2006)
28. Zona marítima especialmente sensible. Secretaría General de Transporte. Dirección General de la Marina Mercante. Capitanía Marítima en Las Palmas. Ministerio de Fomento. Gobierno de España (2007)
29. Lilliefors, H.: On the Kolmogorov-Smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association 62, 399–402 (1967)
30. IALA Recommendation O-134 on the IALA Risk Management tool for ports and restricted waterways (2009)