A CRITIQUE OF THE USE OF DOMAIN ANALYSIS FOR SPATIAL COLLISION RISK ASSESSMENT

Andrew Rawson[a*] and Mario Brito[b]

[a] Electronics and Computer Science, University of Southampton, SO17 1BJ, UK (A.Rawson@soton.ac.uk) – Corresponding Author

[b] Decision Analytics and Risk, Southampton Business School, University of Southampton, SO17 1BJ, UK (M.P.Brito@soton.ac.uk)

ABSTRACT

Predicting the likelihood of maritime accidents is hindered by the relative sparsity of collisions on which to develop risk models. Therefore, significant research has investigated the capability of non-accident situations, near misses and encounters between vessels as a surrogate indicator of collision risk. Whilst many studies have developed ship domain concepts, few have considered the practical considerations of implementing this method to characterise navigational risk between waterways and scenarios. In order to address this, within this paper we implement and evaluate the capability and validity of domain analysis to characterise and predict the likelihood of ship collisions. Our results suggest that the strength of the relationship between collisions and encounters is varied both between vessel types and the spatial scale of assessment. In addition, we demonstrate some key practical considerations in utilising domain analysis to predict the change in collision risk, through a hypothetical wind farm. The outcomes of this study provide research direction for practical applications of domain analysis on collision risk assessments.

KEYWORDS

Ship domain, Collision Risk Assessment, Automatic Identification System

1 INTRODUCTION

Tools to assess and predict the likelihood or consequences of vessel accidents have been key focus of many authors. Comprehensive reviews of these methods have been provided by many (Li et al. 2012, Xu and Wang 2014, Chen et al. 2019, Kulkarni et al. 2020) and include statistical derivation of incident rates (Bye and Almklov, 2019), analytical models (Pedersen, 1995; Mazaheri and Ylitalo, 2010; Li et al. 2012), or the use of Bayesian Networks (Hanninen, 2014). Assessing the accuracy of these models is challenged by the infrequency at which accidents occur relative to the volume of traffic within an area. For example, even in busy areas such as the Dover Straits, the annual collision rate is approximately 1.2 incidents per year (MAIB, 2014). Given that collisions are a leading cause of ship casualties, accounting for 26% of incidents (EMSA, 2019) and potentially resulting in loss of life and pollution, alternative methods have been proposed to assess collision risk.

To overcome this, many authors have proposed the use of non-critical accident situations, near misses or encounters as a proxy measure for vessel risk (Du et al., 2020). As relates to the risk of collision, where two navigating vessels come into contact with one another, a widely adopted approach is that of the ship domain. This concept represents the surrounding effective waters which a navigator wishes to keep clear of other ships or fixed objects (Goodwin, 1975). Where the domain of a vessel is violated, a threat to navigational safety has occurred (Pietrzykowski and Uriasz, 2009). These domains can range from simple circular buffers to more complex, segmented and dynamic shapes (Fiskin et al. 2020). The size and shape can change depending on several factors including the physical characteristics of the vessels, the encounter situation (head-on, crossing or overtaking as defined by the Collision Regulations), manoeuvrability, the human element, metocean conditions and fairway characteristics. Fiskin et al. (2020) provide a particularly good comparison of the different factors taken into consideration in various domain models. Whilst the terminologies used in previous work vary (Du et al., 2020), the underlying premise is that for a collision to occur, two vessels must meet, and by determining where two vessels meet more often, the likelihood of a collision increases.

A ship domain can be approached in one of two ways. Firstly, as a tactical tool for collision avoidance, whereby a minimum domain is desired and therefore an intelligent navigation system might seek to maintain such passing distance when interacting with other vessels (see for example Tam et al. 2009; Im and Luong, 2019). Secondly, as a strategic tool for collision risk assessments, whereby the frequency of domain interactions is used as a proxy measure for collision risk. This enables quantitative analysis of collision risk, supporting a proactive approach to risk management.

This latter concept has parallels to the Heinrich (1931) accident pyramid, similar accident chains have varying levels of severity. By identifying and preventing lesser incidents, such as unsafe acts, more serious incidents such as fatalities could be avoided. He postulated that for each 300 unsafe acts, 29 minor injuries and one major injury or fatality would occur. Within the concept of vessel collisions, the same causal events which lead to domain encounters, would lead to collisions, and therefore determining the frequency of encounters can inform collision risk.

Within the academic literature, a plethora of domain shapes and sizes has been proposed (Wang et al. 2009; Pietrzykowski and Uriasz, 2009; Szlapczynski and Szlapczynska, 2017; Bakdi et al. 2019; Fiskin et al. 2020). Yet, there are relatively few examples of the application of domain analysis to characterise collision risk across waterways (Szlapczynski and Szlapczynska, 2017), or more generally some have criticised the lack of practical applications of maritime risk models (Kulkarni et al. 2020). It is seemingly an accepted logic that where more domain encounters occur, more collisions should be expected, yet few studies have tested this link (Zhang et al. 2016; Du et al. 2020). Furthermore, several authors have challenged the predictive capability of these models (Goerlandt and Kujala, 2014), having tested the reliability and validity of different approaches. Indeed, the significant diversity of methods proposed can be seen as evidence that there is no single ship domain that can be applied to all circumstances, and a diversity of results should therefore be expected. In addition, most studies have been conducted limited to small areas with homogenous characteristics, such as individual ports (Fang et al. 2019) or isolated waterways (Fujii and Tanaka, 1971).

These issues are not purely academic, with ship domain theory commonly applied to maritime risk assessments in applied settings. Within this context, by comparing the base case number of encounters with an altered future case, the relative difference represents a change in the risk of collision. Such an outcome can then be used to inform consenting decisions by regulators or the cost benefit of an expensive risk control. For example, ship domains have been used to assess the navigational risk implications of infrastructure such as offshore wind farms (Marico Marine, 2018) or oil and gas terminals (van Dorp et al., 2008; 2014). These navigation risk assessments (NRAs) are crucial in ensuring decision makers are able to identify, assess and evaluate maritime risk correctly. Given the outstanding research questions related to the use of ship domains for NRAs, there is a need to resolve these issues to avoid applications providing either erroneous or misleading results.

Within this paper we provide a critique of the use of domain analysis to assess navigational risk. We implement a simple domain at a national scale to demonstrate how collision risk could be characterised using this approach. From this, we then test the statistical relationship between the number of encounters and the historical number of collisions, to consider the validity of such an approach. Finally, we extend the analysis to consider how domain analysis could be implemented to predict the change in collision risk following a new development.

This work offers a number of key contributions. Firstly, we test the relationship between encounters and historical collisions using a statistical analysis at a national level. We demonstrate that there are few studies which have used domain analysis for this purpose, and fewer have evaluated the validity of this approach. Secondly, by utilising a national approach, using a significant volume of data, we are able to offer a more generalised approach than previous studies within specific waterways which have their own local conditions. Thirdly, we contribute a framework and case study through which this approach can be utilised to predict the change in collision risk following a development. Our work demonstrates some clear limitations which warrant further investigation.

This paper is structured as follows; Section 2 describes the key literature for domain analysis and collision risk. Section 3 describes the datasets, methodology and model implemented within this project. Section 4 describes the results of the analysis. A discussion is contained in Section 5 to consider the key findings of this assessment.

2 LITERATURE REVIEW

The statistical and geometric models widely utilising in the literature are static representations of flows of traffic and are therefore limited in their ability to represent the complexity of the maritime situation (Mazaheri et al. 2013). The increased capabilities of computing and increasing availability of vessel traffic data have allowed for a parallel stream of temporal-spatial modelling. Unlike geometric analysis, which characterises traffic as linear flows, the construction of simulations allows for the reflection of the dynamic nature of vessel behaviour. Through this approach, collision risk can be assessed by comparing the time and distance to closest point of approach or by constructing vessel domains and evaluating interactions and encounters between those domains. Amongst collision avoidance and near miss detection/trajectory processing, waterway risk assessment is one of the key applications of domain analysis (Szlapczynski and Szlapczynska, 2017). Whilst innovative new methods are continual proposed, such as the use of Convolutional Neural Networks and artificial intelligence to evaluate traffic situations (Zhang et al. 2020), ship domain analysis has achieved widespread adoption. Within this section we review the previous work on both domain modelling in collision risk assessment and the evaluation of maritime risk models.

2.1 DOMAIN MODELLING IN COLLISION RISK ASSESSMENT

The general rationale of a domain analysis approach takes the form as Equation 1, where N_c is the number of encounters and P_c is the causation probability that an encounter becomes a collision (Goerlandt and Kujala, 2014).

$$f = N_c P_c \tag{1}$$

Therefore, in order to predict the collision risk, we must first implement a domain model that counts the number of encounters between vessels within an area. Secondly, we must determine a conditional probability that an encounter becomes a collision. This latter approach could be achieved statistically, by comparing the number of encounters per unit time with the historical incident record, or by utilising a Bayesian Network to include a variety of contributory factors. However, it is evident that a significant emphasis is placed on the number of encounters within this approach. Such approaches have been presented by several authors, but practical applications of domain analysis are generally limited within the academic literature (Rawson et al. 2014; Szlapczynski and Szlapczynska, 2017).

Qu et al. (2011) present collision risk per leg in the Straits of Singapore by applying fuzzy domain analysis, in combination with measures of speed profiles. The results allow determination of which legs have higher collision risk profiles than others, but this is not compared to historical incident rates. Other approaches are taken by Fang et al. (2019) for Xiamen Port and Feng et al. (2019) in Yangtze Port. Whilst the methods vary, in both these cases, encounters are used to identify where collisions are more likely, spatially, across a port's waterway. These studies are isolated to single specific waterways, lacking generalised applicability to large areas.

Rong et al. (2019) used domain analysis to evaluate near collision scenarios off the Portuguese Coast using three months of vessel traffic. Their use of the Fujii and Tanaka (1971) method with a collision risk evaluation, identified 1,671 near collision scenarios. Whilst the results show the spatial variation in where collision risk is perceived to be highest, these results are not evaluated or validated against any other measures of collision risk. Zhang et al. (2016) conducted near miss modelling in the Gulf of Finland, using k-means clustering to determine types of encounters.

Rawson et al. (2014) presents a domain model applied for collision risk on the River Thames, developed through consultation with local skippers. In this case, not only is the spatial variation in encounters used to determine where collisions are more likely, the relationship between traffic volume and number of encounters is investigated to predict how collision risk might change with increased numbers of vessels navigating on the future. Unlike many of the studies described above, this sought to develop domain analysis from a descriptive to a predictive tool.

The motivation to predict changes in collision risk have led to the inclusion of domain analysis in a number of major applied studies. For example, in van Dorp et al. (2008; 2014), a simulation is developed of vessel traffic in the waters of Washington State, USA. Collision candidates are determined every minute of the simulation based on proximity, and a probabilistic causation probability assigned based on expert judgement elicitation. Upon this, the model results are compared having included new traffic associated with a proposed marine terminal, and then various risk control measures are implemented, and the benefits compared. Domain analysis in this context is a decision support tool to determine both the acceptability of the risks associated with the development but assess the cost benefit of risk controls such as escort towage.

Other practical implementations include the aforementioned NRAs for offshore wind farms in the UK, where the baseline and future case risks associated with the development can be compared (Marico Marine, 2018; Anatec, 2019). The relative change in encounters informs decision makers on the potential increase in risk associated with a development.

2.2 VERIFICATION OF MODEL RESULTS

An interesting aspect which is often omitted from studies on domain analysis, and is generally assumed, is how representative domain analysis results are for either characterising or predicting collision risk. Within their review, Du et al. (2020) note that there is comparatively little academic research focusing on the validity of non-accident critical event detection models. This is both a result of the relative sparsity of accidents upon which to validate the model, and the complexity of the accident scenario due to inter-relating contributory factors.

Aven and Heide (2009) define model validity as the degree to which the risk analysis describes the specific concepts that one is attempting to describe, including amongst others, the degree to which the model risk scores reflect the underlying true risk. A significant contribution to framing this discussion is that of Goerlandt and Kujala (2014). They evaluated the reliability and validity of collision risk models in a Traffic Separation Scheme (TSS). Their results showed that by using different models, significantly different frequencies and spatial distributions of collision risk were obtained. This raises questions on the predictive capability of domain analysis for collision risk assessment. This can be further evidenced by the wide variety of models presented within the literature (Goerlandt and Kujala, 2014), with different shapes, functions regarding speed or encounter type and numerous other variables.

By way of example, the application for an offshore wind farm in the UK was assessed by two consultancies using two forms of domain analysis in order to measure the relative impact of the development on collision risk (Marico Marine, 2018; Anatec, 2019). One assessment presented an increase in collision risk of 50% from approximately once in six to approximately once in four years whilst the other reports a change from once in 47 years to once in 46 years, a 2.2% increase. These differences are significant, demonstrating that a great variability in results can be derived by alternative implementations of the same methodological approach.

One method to validate the results of an assessment might be to compare the model results with historical accident numbers, such that where most domain encounters occur, is where most accidents have occurred. Where accident data is unavailable, the validity of derived high encounter intensity regions can be compared against perceived high-risk regions by consulting with local practitioners. For example, in Rawson et al. (2014), the results were recognisable to local stakeholders, however, this does not necessarily support the underlying statistical validity. A further method would be to evaluate the method against some other method, for example, Zhang et al. (2016) compare their encounter results against previous studies in the same area but using different methods.

Within this paper we seek to test this relationship and provide some research direction on the application of domain analysis for collision risk assessments. We ask two key questions; firstly, what is the validity in the implementation of domain analysis to characterise maritime risk spatially? We achieve this by analysing domain encounters across the continental United States during June 2018. Secondly, what are the research challenges in utilising this approach to predict the change in risk following future changes in traffic activities? In order to test this second aspect, we utilise a hypothetical case study of a new development in the Strait of Juan de Fuca, Washington State.

3 METHODOLOGY

Within this section, the methodological approach is described, including the underlying datasets, utilised domain function and statistical method to determine the relationship between encounters and collisions. In addition, we present a framework for utilising encounters to assess the impact of future scenarios.

3.1 DATASETS

3.1.1 Vessel Traffic Data

Vessel traffic data from the Automatic Identification System (AIS) was used to model the navigation of vessels within the study area. AIS is a transponder system required to be fitted to all commercial vessels, under SOLAS Chapter V, and voluntarily carried by smaller craft, that sends and receives information about the dynamic movement (position, speed, course and others) and static identification (type, size, name and others) of navigating vessels (IALA, 2002). Whilst the system was principally developed for improving maritime safety, AIS data enables high resolution analysis across a wide range of applications (Fournier et al. 2018; Yang et al. 2019).

The Marine Cadastre (2020) publish AIS data collected the US Coast Guard’s national network of AIS receivers. Data was extracted from the 1st June to 30th June 2018, containing 32GB of AIS data in csv format. The dataset was processed to filter to the study area and linear interpolation used to standardise the dataset to one position per vessel per minute, provided the time between sequential positions of the same vessel did not exceed ten minutes, whereby it is assumed that tracking of the vessel has been lost and the interpolated positions have insufficient accuracy for use in collision risk assessment.

As AIS often contains missing or erroneous descriptions of vessels (Harati-Mokhtari et al. 2007), additional datasets were connected to that vessel containing summaries such as vessel type and length. Missing values were either filled in or if no vessel type information was found, the vessel was omitted from the final dataset. The resulting vessel traffic dataset contained approximately 120 million vessel positions and a density plot of the collected data is shown in Figure 1.

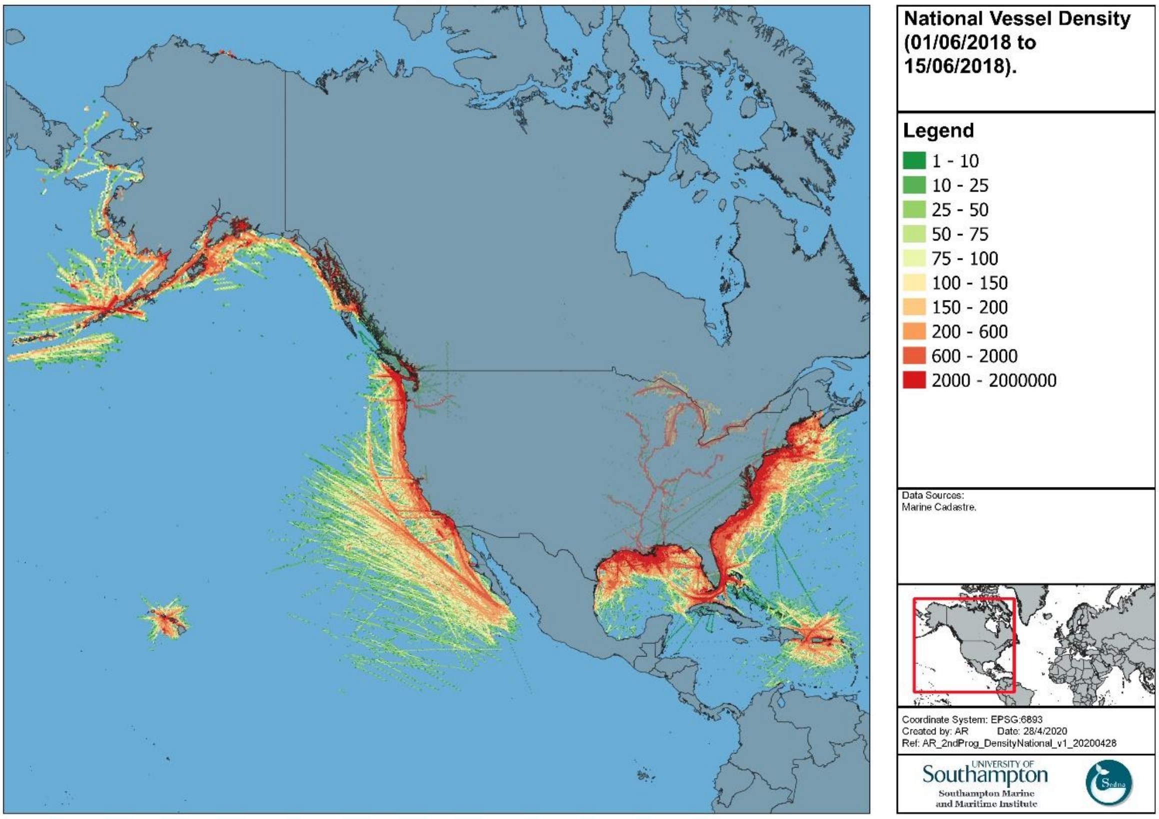


Figure 1: Vessel traffic density.

3.1.2 Incident Data

Under the Code of Federal Regulations 46 CFR 4.03 / 4.05, any marine casualty or accident occurring within the United States navigable waters, including grounding, collision, allision or flooding, shall be reported to the Coast Guard. A database of these incidents from 2002 to July 2015 is available specifically for use by researchers (USCG, 2020).

Two key files are relevant, the MisleVslEvents contains 132,717 entries and 20 fields for incidents involving vessels and the MisleVessel contains the details for 1.35 million vessels with 66 fields. Both datasets contain a vessel_id field which enables joining ship attribute data and incident data. The dataset was filtered to collisions only, often leaving two records per incident (signifying two or more vessels involved in a collision). In total, 5,165 instances of vessels involved in a collision across 2,524 unique collisions were used, with the locations of collisions shown in Figure 2. Most of the collisions are located at the locations of some of the country's major ports and along the Mississippi River.

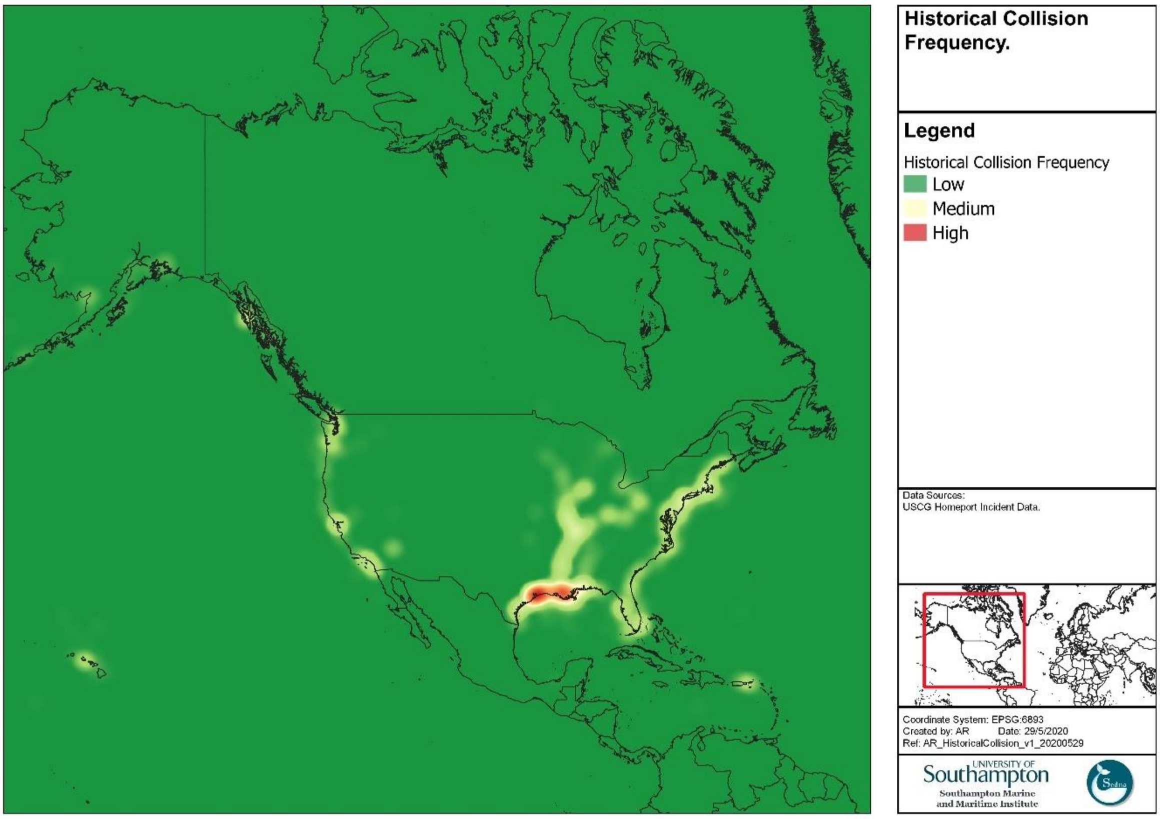


Figure 2: Heatmap of historical collisions in US.

3.2 DOMAIN MODEL

There are a significant number of different domain models from which to choose for a given study. Given the advantages discussed below, we choose to implement the domain model proposed by Wang (2010), with some modifications, within this section. Firstly, based on the reviews of Szlapczynski and Szlapczynska (2017), and Fiskin et al. (2020), we note that dynamic domains which change based on conditions, most commonly vessel length and speed, are popular amongst different authors and therefore our chosen domain model should reflect this. Furthermore, we do not wish to include many other factors as this increases the computational complexity, particularly given the significant scale of our assessment. Secondly, we wish to utilise a domain model which has been widely discussed in the literature and there are other studies utilising this approach, such as Qu et al. (2012), and Goerlandt and Kujala (2014). Thirdly, the model must be well documented and therefore reproducible.

Several authors, including Wang (2010), have proposed non-binary collision models whereby the significance of encounters is graduated. By contrast, binary domain models consider that two vessels either encounter or do not encounter, given the rules of the domain, and therefore one encounter is equal to another. One method to achieve a non-binary domain is to utilise fuzzy boundaries, whereby vessels that are outside of the fuzzy boundary are not at risk, but as the vessel is closer inside the fuzzy boundary the risk increases (Pietrzykowski and Uriasz, 2009). Another approach is to apply what is often referred to as a Vessel Conflict Ranking Operator (VCRO) which grades the severity of different types of encounter, for example, Zhang et al. (2016). The use of non-binary domains is useful in situations of collision avoidance, such that as two vessels come closer together, the risk of a collision increases and therefore the need for a response is heightened. However, in terms of strategic collision assessments, few studies have suggested that where more significant encounters occur the likelihood of a collision increases. Indeed, it could be argued that a properly calibrated dynamic domain should expand or contract such that the domain boundary is of equivalent risk between different situations. Therefore, we have chosen not to implement a fuzzy or non-binary ship domain, in order to maintain simplicity and efficient computation in this comparative model, however we discuss the implications of this decision in Section 5.

In Wang (2010), the author proposes an Intelligent Spatial Collision Risk Model based on the Quaternion ship domain. The domain is an ellipse shape with four directional radius R_{fore}, R_{aft}, R_{port}, R_{starb} (see Figure 3 and Figure 4). The size of these radius are given by the formulas:

$$\begin{cases} R_{fore} = \left(1 + 1.34\sqrt{k_{AD}^2 + (k_{DT})/2)^2}\right)L \\ R_{aft} = \left(1 + 0.67\sqrt{k_{AD}^2 + (k_{DT})/2)^2}\right)L \\ R_{starb} = (0.2 + k_{DT})L \\ R_{port} = (0.2 + 0.75k_{DT})L \end{cases} \tag{2}$$

Where L is the ship length and k_{AD} and k_{DT} represent the advance and tactical diameter of the vessel scaled by some factor k that represents risk appetite (defaulted to 1), and where V_{own} is the ship speed in knots, as given by:

$$\begin{cases} k_{AD} = \dfrac{A_D}{L} = 10^{0.3591 \lg V_{own} + 0.0952} \\ k_{DT} = \dfrac{D_T}{L} = 10^{0.5441 \lg V_{own} - 0.0795} \end{cases} \tag{3}$$

To calculate whether a target vessel falls with own vessel domain, a pairwise distance calculation is conducted at every timestamp, with distance between vessels i and j with coordinates X and Y (in UTM metres) is derived by:

$$D = \sqrt{(X_j - X_i)^2 + \ (Y_j - Y_i)^2} \tag{4}$$

The identify function as to whether a target vessel falls within the domain can then be described as (Zhang et al. 2016):

$$l_\alpha = \left(\frac{1 + tan^2\alpha}{\dfrac{1}{S^2} + \dfrac{tan^2\alpha}{R^2}} \right) \tag{5}$$

Where l_α is the distance to domain boundary at angle α, S is the R_{port} and R_{starb} radius functions and R is the R_{fore}, R_{aft} functions. Therefore, a target vessel at relative bearing α from own vessel is within the domain if the distance D is greater than l_α.

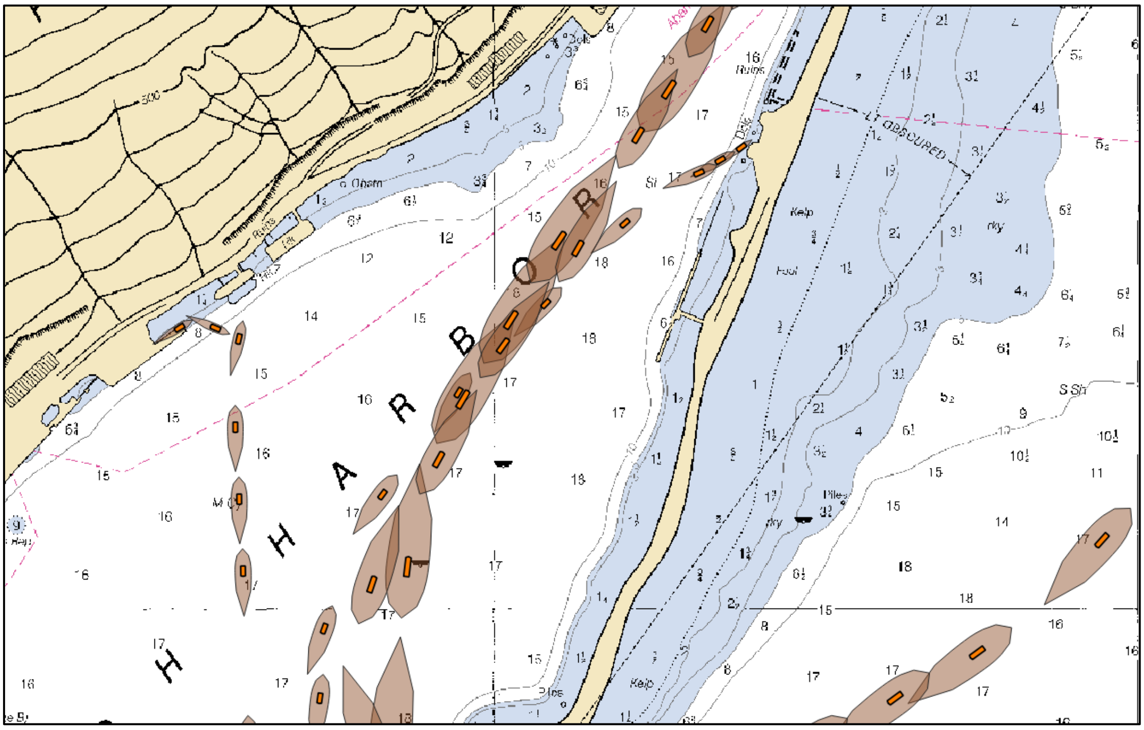


Figure 3: Example implementation of domain model (domains approximated with 18 vertices rather than true ovals).

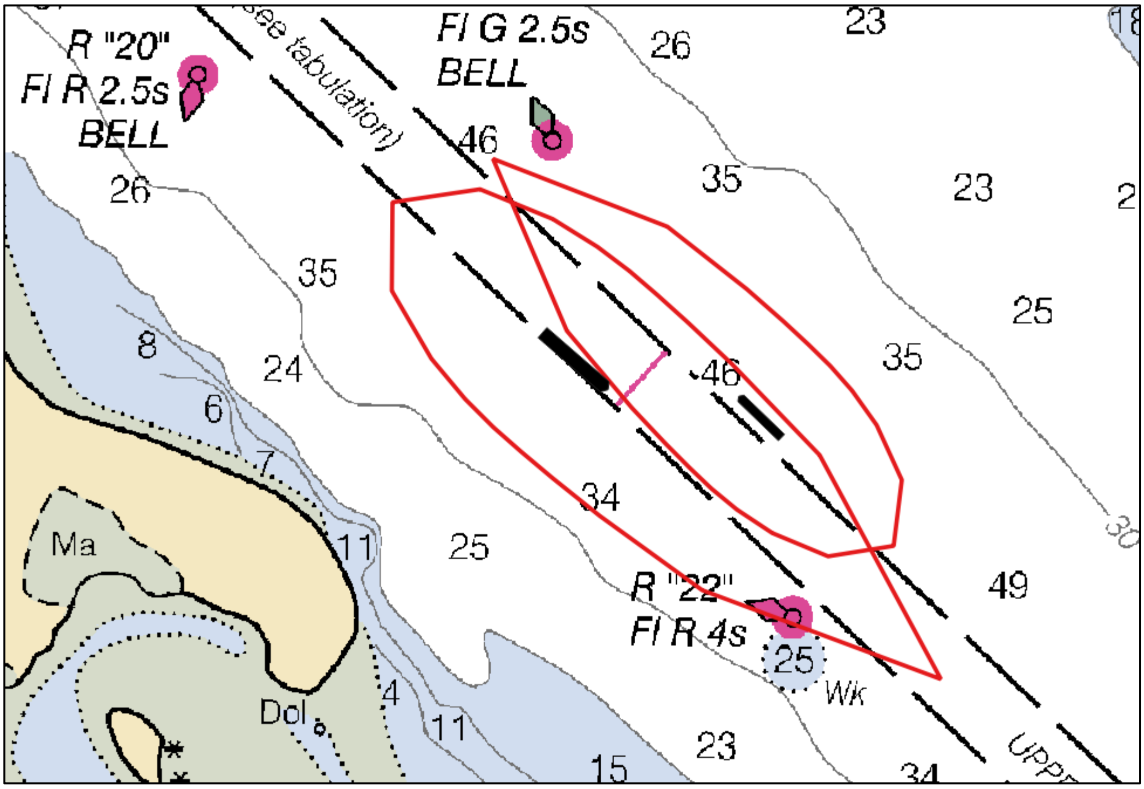


Figure 4:Example encounter in Columbia Channel of an inbound 218m vessel and an outbound of a 142m vessel.

3.3 MEASURING ENCOUNTERS

In order to determine whether an encounter has occurred, the following workflow was developed, which is shown in Figure 5. Within the first part, a pairwise join of the processed AIS data is conducted such that at every minute, all vessels positions are compared. If the haversine distance between vessels is less than the calculated domain distance (described in Section 3.2) and the average speed of vessels is greater than 1 knot, an encounter is retained. This process removes stationary vessels such as those alongside that are not navigationally significant. Unlike previous studies which are limited to singular waterways and therefore able to use Euclidian Distances with metres based projection systems, our analysis covers a significant study area both latitudinally and longitudinally and therefore we are required to utilise haversine distances with the latitudes and longitudes of the vessel positions. The resulting encounter table contains numerous key characteristics about that encounter situation that are further analysed. This stage required an average of 2,800 vessels to be compared every minute for the full month of data and was therefore computationally expensive to run.

A second stage was necessary to process chains of consecutive encounters (i.e. occurring successively every minute during an overtaking situations), which we call Prolonged Encounters. We consider that a single record should be retained for each meeting situation between two vessels, and therefore a method to filter prolonged encounters is required. The encounter dataset is sorted by vessel MMSI numbers and encounter time, and if the encounter has the same vessel ids and occurred less than 10 minutes previously, the Prolonged Encounter ID number (PID) is kept the same. A 10-minute filter was utilised as occasional breaks between successive encounters occurred, before re-establishing the same chain. Finally, duplicate PID values were removed, retaining the encounter in each prolonged encounter that had the minimum separation between vessels.

For example, where two vessels encounter one another at a single timestamp, a single PID value is assigned which is unique in the dataset. If those two vessels were to encounter successively for five minutes, five encounter records are recorded, but each of those five records shares the same PID number. By removing duplicates, only the closest position of those five is retained, and similarly the single record in the first example is retained. The resulting dataset contained 348,543 encounters across the United States during June 2018.

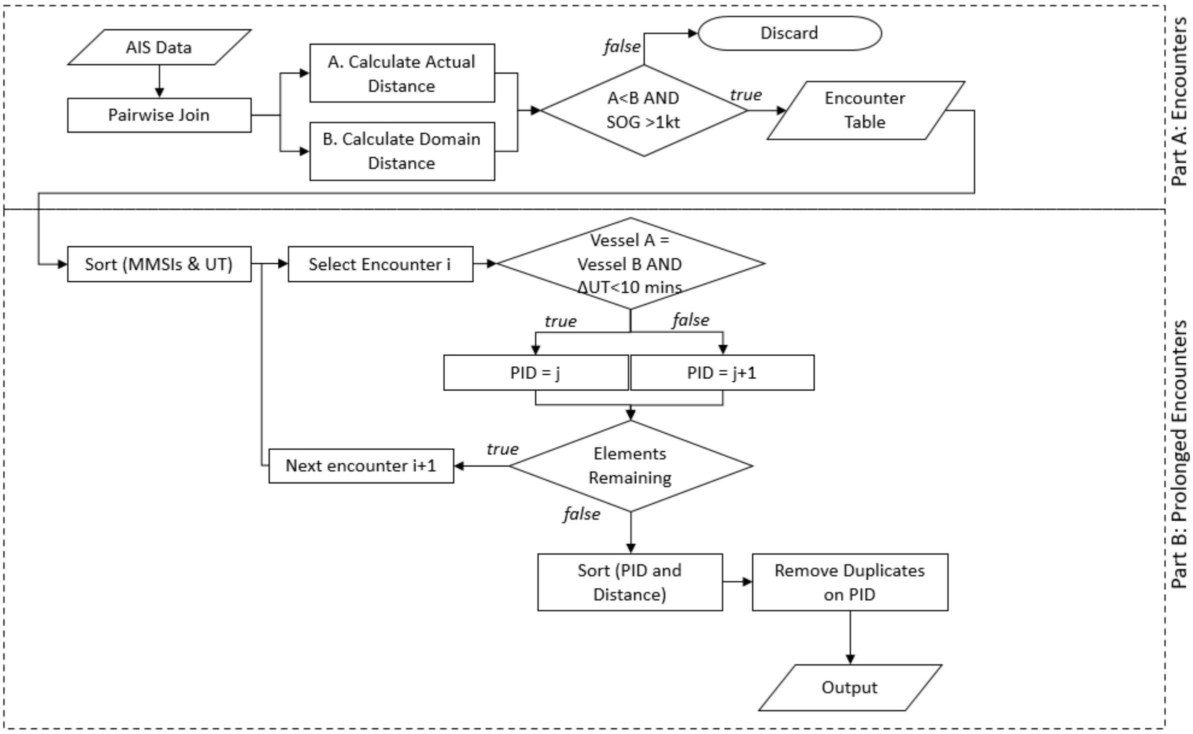


Figure 5: Encounter Processing Flowchart.

3.4 SPATIAL MODEL OF ENCOUNTERS AND INCIDENTS

In order to analyse the results and calculate correlations, a spatial data structure is required into which the data is aggregated. One method would be through cartesian grids, however many authors have described inherent limitations of these formats when performing spatial analysis, such as distortion of shape, non-uniform adjacency between cells and variable cell areas when implemented at global or regional scales (Sahr and White, 1998). To overcome this, Discrete Global Grid Systems (DGGS) have been proposed, which partition the world into equal area platonic solids, such as hexagons or triangles (Sahr et al., 2003). In this case, we utilise hexagonal grids which have been shown to exhibit certain inherent advantages, such as uniform adjacency between cells and visual interpretability (Birch et al., 2007).

As part of the SEDNA project, the University of Southampton developed a python library that implemented the DGGRID R library (Barnes, 2016), called dggridpy (Correndo, 2019). The package enabled DGGS cells to be constructed at varying resolutions, and spatial data indexed. A hexagonal ISEA aperture 4 DGGS was developed at resolution 7, 8, 9 and 10 to support the analysis at different scales. Table 1 and Figure 6 compares the four resolutions assessed in this paper.

Table 1: Details of DGGS Resolutions

Resolution	Number of Cells (Global)	Cell Area (km^2)	Centroid Spacing (km)
7	21,872	23,322	151.2
8	65,612	7,774	82.3
9	196,832	2,591	50.4
10	590,492	863	27.3

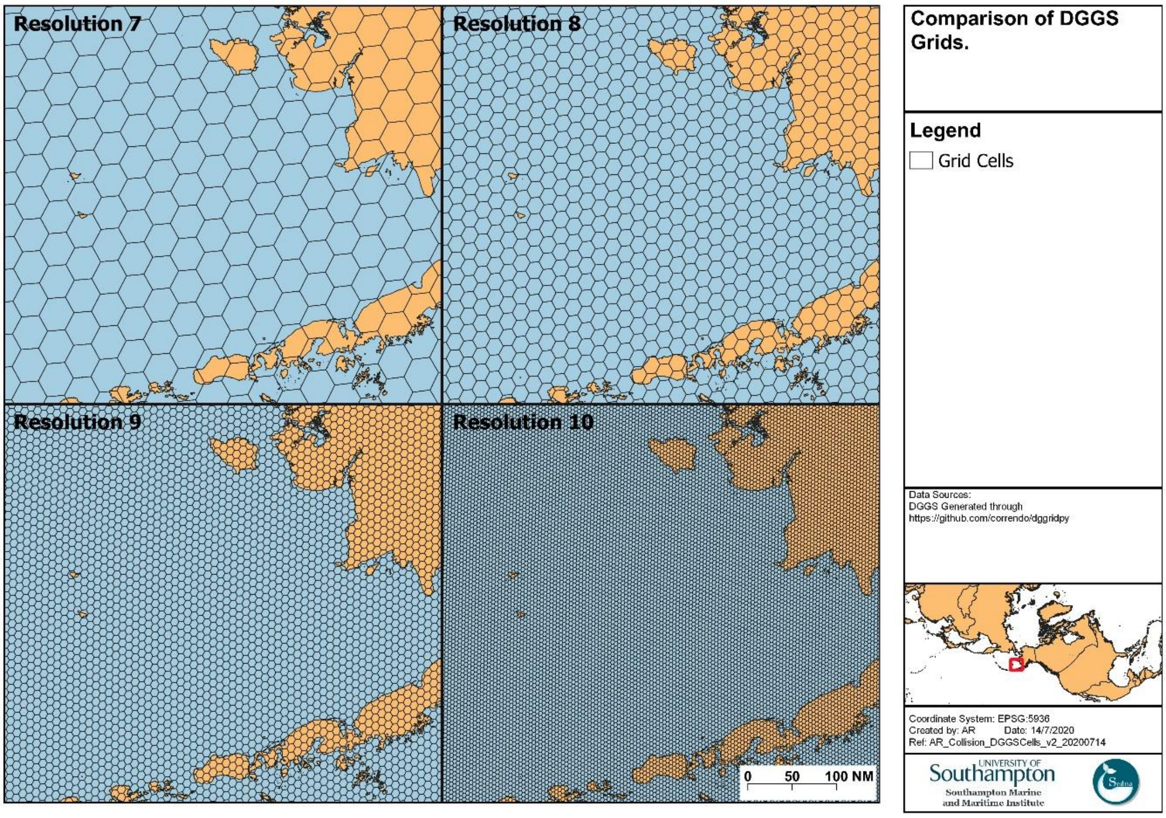


Figure 6: Comparison of DGGS Grids across Alaska.

3.5 FUTURE SCENARIO MODELLING

Whilst rarely considered in the literature, a key potential application of domain analysis is predicting the impact on collision risk of some future change in traffic. This might include an increase in the number of vessels, or alternatively, as taken in our example here, some future development, such as an offshore wind farm or new routeing measure. Such a development would require vessel traffic to divert around an obstacle, potentially creating choke points and increasing collision risk. Were the wind farm located elsewhere, with different geometries or traffic flows, the change in collision risk could be compared, potentially enabling optimisation of layouts to minimise the impact on collision risk. However, within this study we consider only a single development.

In such a circumstance, we can predict the expected number of collisions by multiplying the encounter frequency by a conditional probability (P_c) that an encounter would result in a collision ($P_{Collision}|P_{Encounter}$). We could measure the difference in collision risk by calculating the number of encounters under both the base and future case scenarios, provided the latter case can be sufficiently modelled.

$$\text{Collisions/Year} = Encounters/Year * \ P_c \tag{6}$$

Figure 7 shows a visualisation of a typical scenario which might be encountered. In this case, some physical obstruction is placed in the flow of traffic, which requires an alteration of a vessel's course to avoid the obstacle. Understanding how this scenario might change collision risk is not a trivial problem, and essential to forecasting the increase of risk of a proposed development or evaluating the requirement for risk control measures such as routeing schemes.

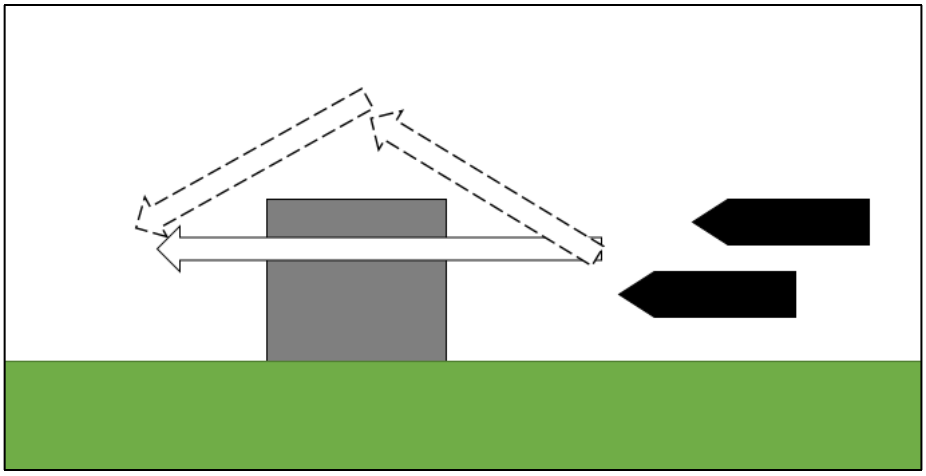


Figure 7: Dummy example of future scenario modelling.

3.5.1 Case Study

To demonstrate this, we propose a hypothetical situation of an offshore wind farm to be constructed in the Straits of Juan de Fuca, Washington State (Figure 8). The Straits are the major approach routes between the continental USA and Vancouver Island, serving several major ports such as Vancouver, Seattle, Everett and Victoria. The route has a TSS, which mandates single direction travel for large commercial vessels. In our situation, the hypothetical wind farm boundary extends to the limits of the southern (inbound) traffic lane.

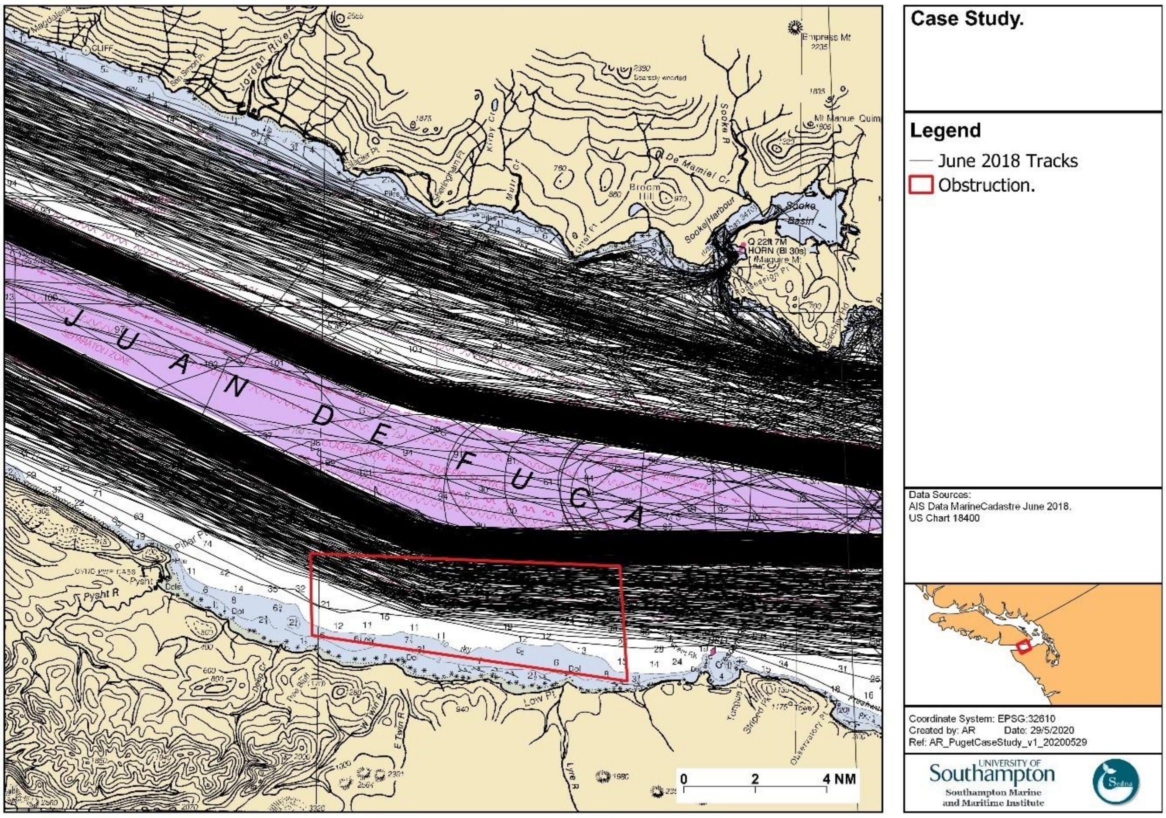


Figure 8: Vessel traffic in the Straits of Juan de Fuca and hypothetical wind farm (June 2018).

In order to test how this development might change collision risk, we must first calculate the baseline number of encounters, as described in the preceding sections. Following this, a method of re-routeing traffic flow is required, and then the domain model re-tested with the new traffic flows to measure the increase.

3.5.2 Re-Routeing Vessel Traffic

Numerous methods are available for routeing vessel traffic, including the use of path finding algorithms such as A* or Dijkstras (Ari et al., 2013). In this case study, a simple diversion method is utilised, which is described in Figure 9 below. The method takes tracks that pass through an obstruction and assigns new coordinates related to boundary markers that indicate the expected course for vessels around the obstruction. The process is to:

1. Identify vessel trips, taken by the same vessel, such that the subsequent positions are less than 10 minutes apart. If not, create new trips.

2. Spatial query as to whether a trip passes through the obstruction area.
    a. If not (i.e. transit A in Figure 9), set aside.
    b. If so:
        i. Determine the closest boundary coordinates to each position using a KDTree nearest neighbour search.
        ii. Drop duplicates for each trip ID and boundary coordinate.
        iii. For each trip, assign a random track offset using a normal distribution with standard deviation 100m.
        iv. Create new coordinates using boundary coordinates and track offsets.
3. Merge diverted and un-diverted positions to new database.

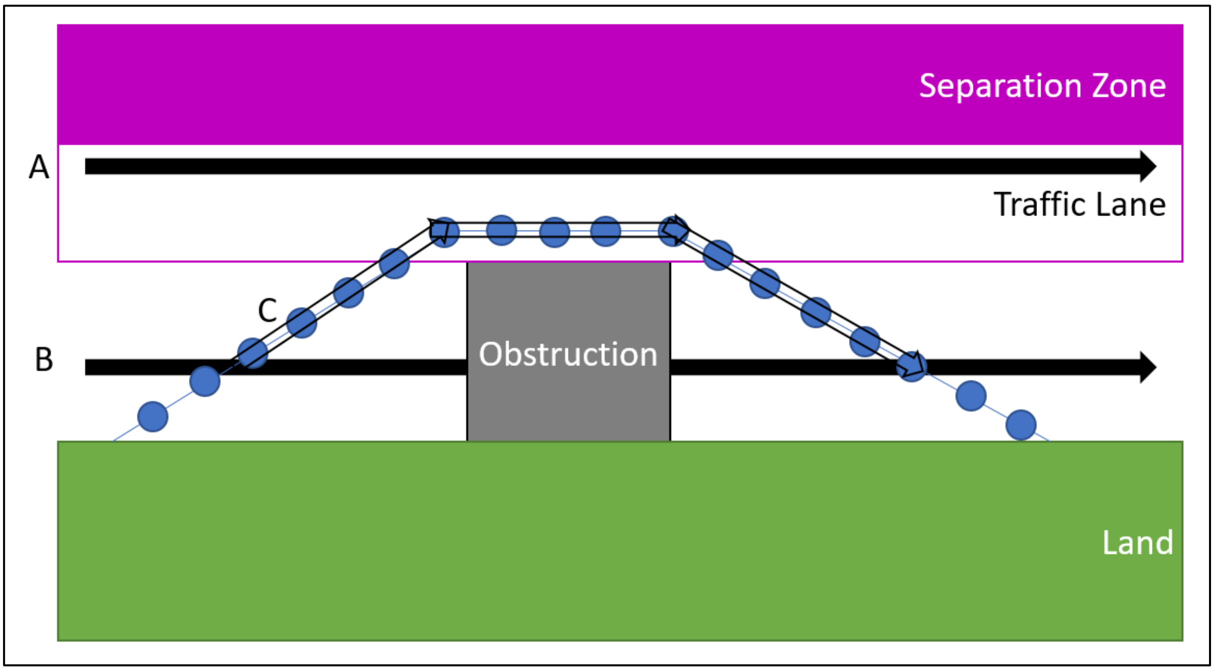


Figure 9: Route diversion method.

Figure 10 shows the results of this on the June 2018 dataset. Several notes should be made on the applicability of this method in real scenarios as opposed to our hypothetical one. Firstly, the method has maintained transit in the traffic lane only, hence there is a significant concentration of traffic to the north of the obstruction which would be unrealistic. In practice, it would be likely that the traffic lanes would be reconfigured to maintain sufficient separation from the obstruction. Secondly, we have assumed that vessels transiting outside of the lanes normally would continue to transit in this fashion except when passing the obstruction. This results in significant alterations of course for some vessels that are close to the southern shoreline. Thirdly, we have not included any alteration of transit time, assuming that vessel speed is increased to cover the increased distance. Given the relatively small increased transit distance in this example, this is not considered a significant limitation, but may be much more pronounced in other applications and therefore warrants further work.

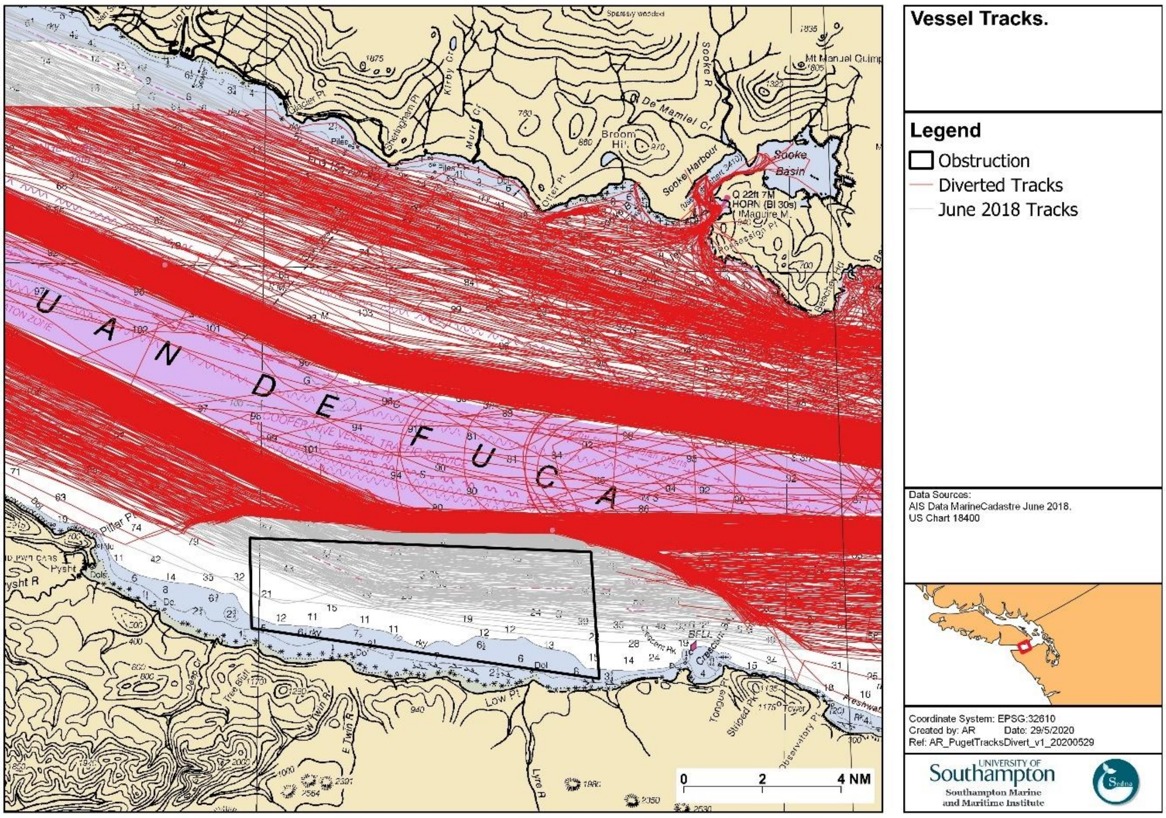


Figure 10: Diverted vessel traffic.

4 RESULTS

4.1 IMPLEMENTATION OF DOMAIN ANALYSIS

The June 2018 AIS data was processed and the number of encounters, scaled to annual figures, in each location mapped in Figure 11 and Figure 12. For much of the county, the number of encounters is low, but where vessel traffic is concentrated, such as the approaches to ports, the frequency is increased. Figure 12 compares the number of encounters and historical number of collisions in the Gulf of Mexico, with the results suggesting a relationship does exist. It is evident that mapping the distribution in encounters can show differences in traffic disposition between locations, but we must then test the strength of the relationship between encounters, and collision frequency.

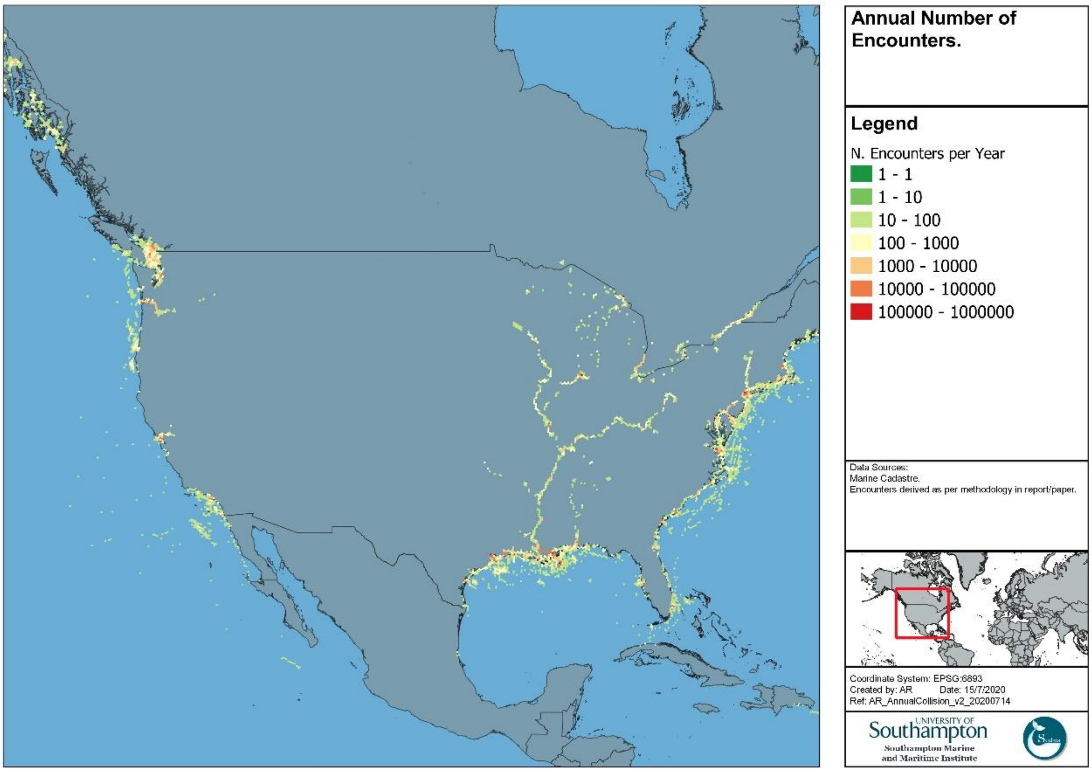


Figure 11: DGGS9 Encounter Rates.

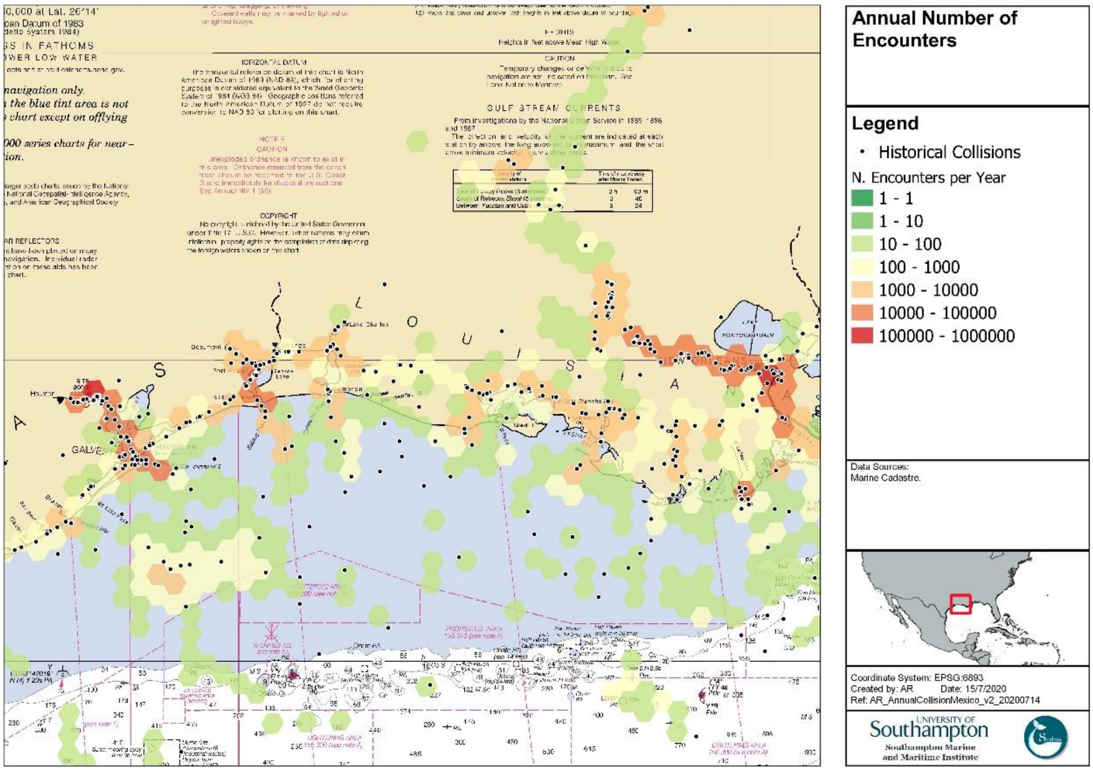


Figure 12: DGGS9 Encounter Rates and Historical Collisions in Gulf of Mexico.

4.2 RELATIONSHIP BETWEEN ENCOUNTERS AND COLLISIONS

The Pearson R (SciPy, 2020) values across a number of spatial grid resolutions and for each vessel type are shown in Table 2. Firstly, in agreement with previous research the results are subject to the Modifiable Areal Unit Problem (MAUP), that results in the correlations increasing as the cell sizes become larger and the results more generalised (Rawson et al. 2019). The resolution at DGGS9 was taken forward as a compromise between high spatial resolution but maintaining good correlation and the relationship at this resolution is shown in Figure 13.

Table 2: Pearson R Values (indicates p value greater than 0.05).*

	Encounters				Exposure	Variants	
	DGGS7	DGGS8	DGGS9	DGGS10	DGGS9	DGGS9 (>1nm)	DGGS9 (>10nm)

Total	0.63	0.57	0.53	0.36	0.66	0.50	0.06
Commercial	0.71	0.65	0.61	0.52	0.33	0.63	0.07
Fishing	0.27	0.22	0.18	0.15	0.18	0.11	0.01
Passenger	0.32	0.32	0.28	0.27	0.38	0.30	0.01
Recreational	0.54	0.51	0.24	0.27	0.41	0.38	0.01*
Other	0.64	0.58	0.54	0.33	0.71	0.24	0.01*

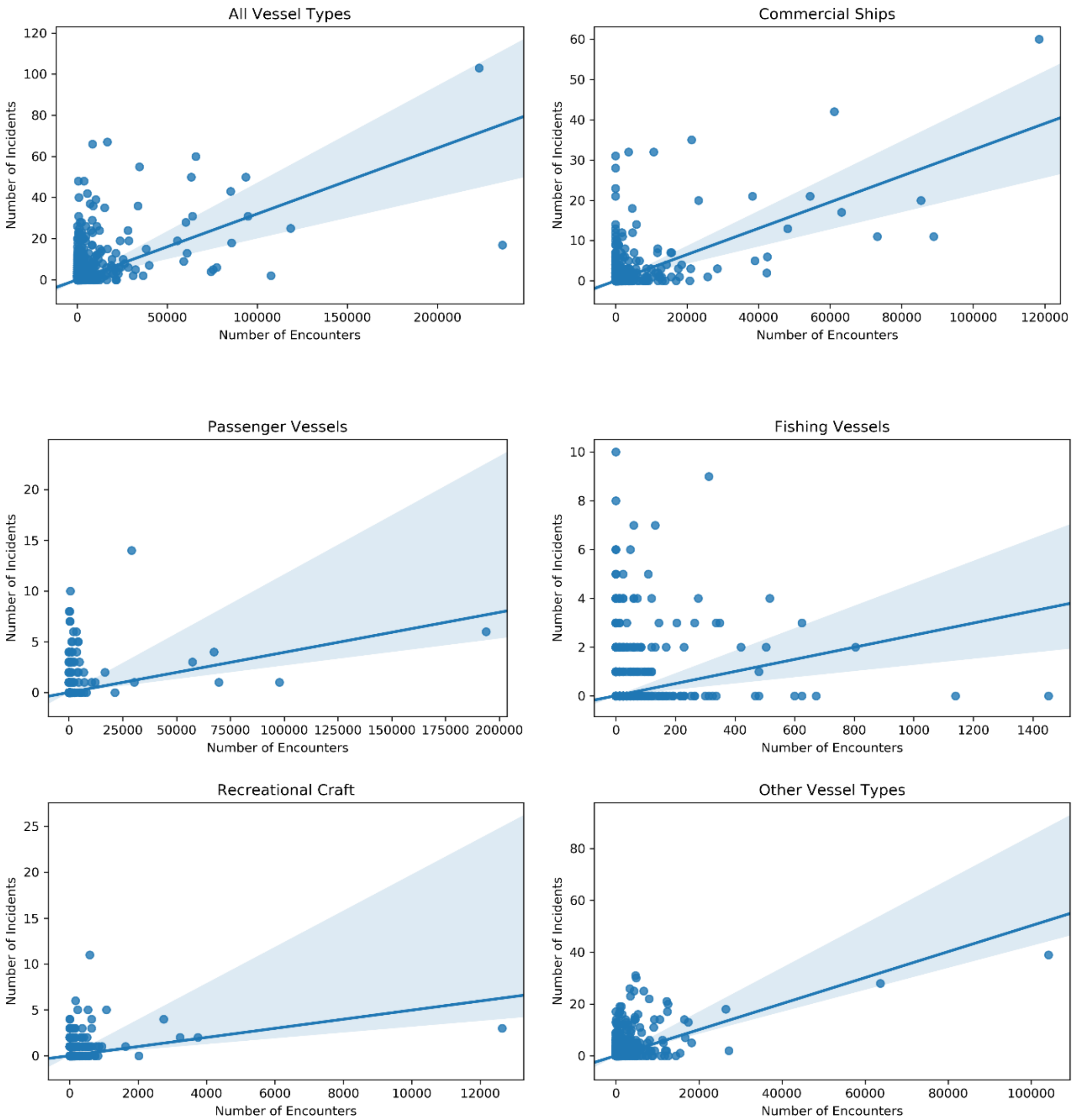


Figure 13: Relationship between encounters and collisions at DGGS Resolution 9.

Secondly, the results are varied between vessel types, with commercial having the greatest at DGGS9 of 0.61 and fishing vessels having the weakest at 0.18. Other vessel types, such as tug and tows and dredgers have similar correlations, exceeding that of passenger and recreational vessel types. The relative low correlation of fishing and recreational maybe partly explained by the low carriage of AIS, which might be disproportionately greater in inshore areas where encounters are more likely to occur than offshore, where traffic is generally less concentrated. In addition, some fishing activities require multiple vessels to operate in close proximity, generating encounters but not necessarily increasing collision risk. There is also a question of incident data completeness, that might result in higher reporting rates for commercial or passenger vessels than fishing or recreational vessels.

Thirdly, the results are compared against a simple count of exposure (hours of operation per grid cell) at the same resolution. Interestingly, for some vessel types this had a higher correlation for historical collisions than the number of encounters. The results suggest that passenger, recreational and other vessel collisions can be better predicted by vessel exposure, although commercial vessel collisions are better described using encounters. This suggests that simple methods of risk assessment, such as using simple measures of traffic exposure, can achieve near equivalent results, without the need for significant computation.

Finally, sensitivity testing was conducted at DGGS9 resolution by including only those grid cells that were more than 1nm and 10nm from the coast, respectively. At 1nm, which includes coastal cells, but excludes inshore cells (such as the Mississippi River), the correlations are largely unchained. However, at 10nm, the correlations reduce significantly. This might be partly due to reduced coverage of AIS at greater distance from shore, reduced sample size of collisions, or indicate a possible difference in navigation behaviour offshore as opposed to coastal. Whilst this latter point is logical, vessels navigating offshore have more space and time to plan avoiding manoeuvres than in constrained waterways, due to the aforementioned limitations we are unable to conclusively demonstrate a difference in navigation behaviour between these settings.

Furthermore, using the analysed data, it is possible to compare the relationship between encounters and exposure (Figure 14). In areas with significant numbers of movements, it would be expected that vessels would inevitably navigate closer together. For most vessel types the results support this, yet fishing vessels show a significant number of low exposure and high encounter locations, likely situations where fishing vessels are working equipment in groups. This further supports the finding that vessel traffic exposure may be equivalent to vessel encounters as a predictor of collision frequency. In Rawson et al. (2014) it was shown that the River Thames exhibited a non-linear relationship, as the river became busier with more activity, the number of encounters increased exponentially. It is likely that different environments will have different relationships between these variables, for example as a function of the waterway geometry, and therefore further research is necessary using different variables or statistical models to better describe this relationship.

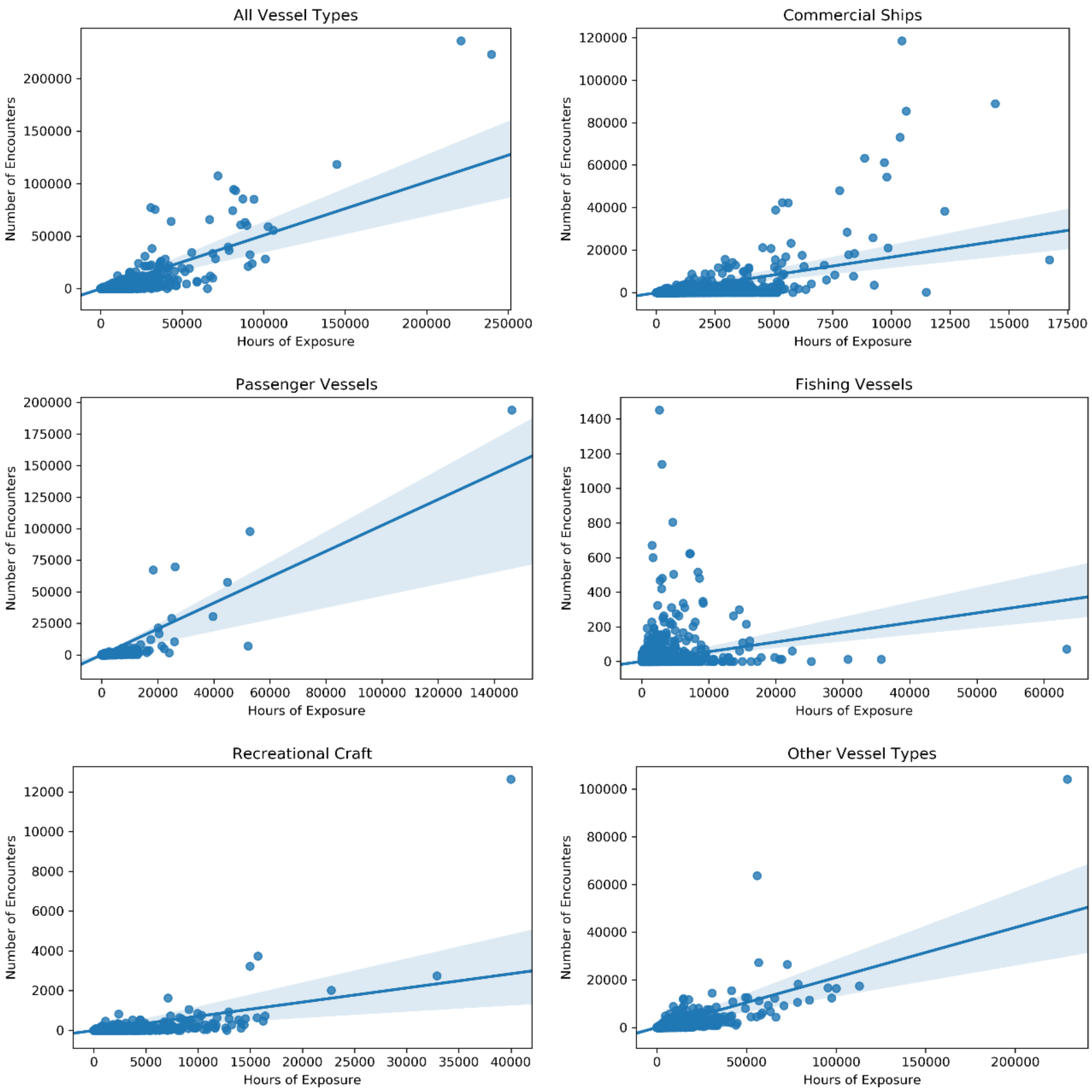


Figure 14: Relationship between encounters and exposure at DGGS Resolution 9.

4.3 FUTURE SCENARIO MODELLING

Finally, the results of the future scenario model are described in this section. Based on the re-routeing of traffic conducted, the number of encounters before and after the development can be established. Table 3 shows these results, with the re-routeing resulting in a 234% increase in the absolute number of encounters, and a 778% increase in unique encounter situations. If we annualise the number of encounters and apply a causation probability of ($P_{Collision} | P_{Encounter}$) where Pc is equal to 4.9×10^{-5} (Goerlandt and Kujala, 2014), then annual collision probabilities can be obtained. The results show an increase from once in 189 years to once in 21.5 years.

These encounters are shown in Figure 15, and are clustered adjacent to the obstruction, where traffic flow is most concentrated. This methodology enables quantification of the magnitude of the impact of the development on collision risk, and presentation of the spatial distribution of that risk, in order to target risk controls effectively. A discussion of some of the key assumptions in this approach is contained in Section 5.3.

Table 3: Results

Model	Encounters (June18)	Unique Encounters (June18)	Unique Annualised	Causation Probability (Pc)	Collisions/ Year
Baseline	166	9	108	4.9×10^{-5}	5.29×10^{-3}
Re-Routed	553	79	948	4.9×10^{-5}	4.65×10^{-2}

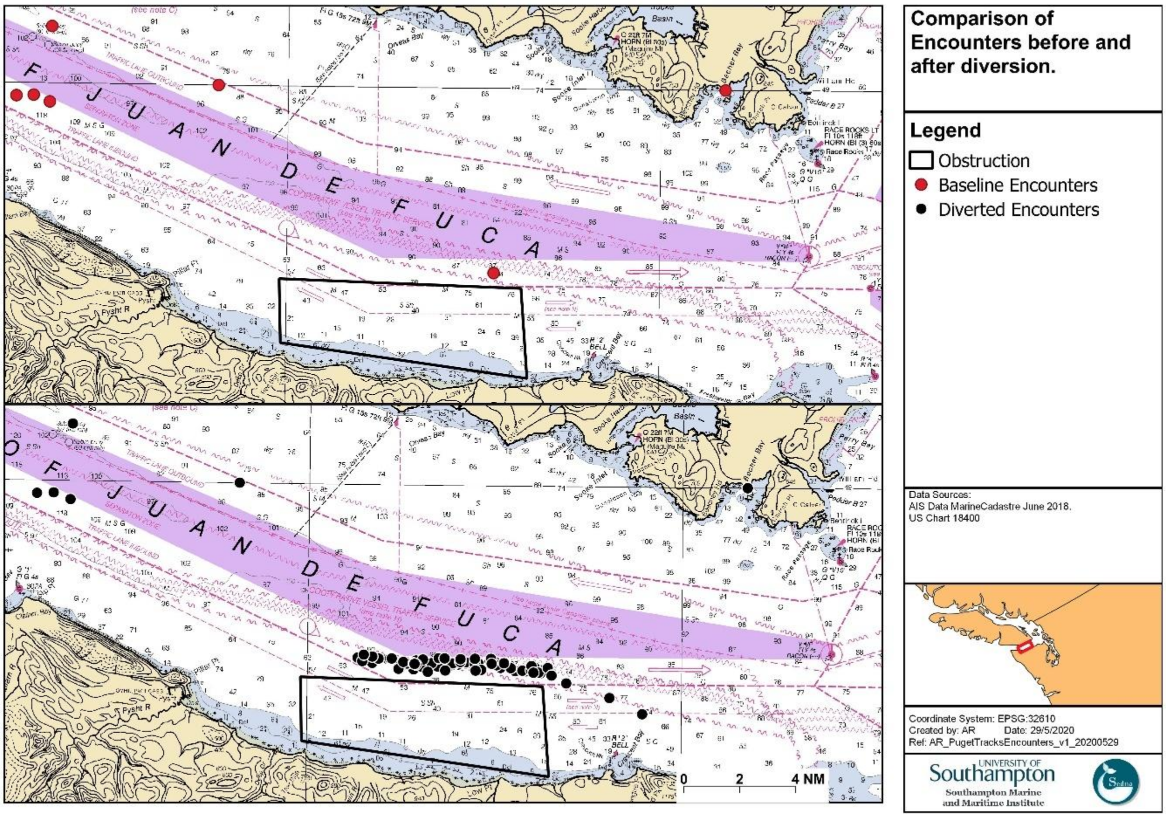


Figure 15: Comparison of encounters between models.

5 DISCUSSION AND FUTURE WORK

Within our discussion we consider three key findings of our work. Firstly, we consider the statistical relationship between the number of encounters and collisions and the implications of this result on collision risk modelling. Secondly, we consider how specific implementations of ship domains might impact the results, and whether some models are more suited to one environment than another. Finally, we consider some of the key methodological aspects of our future scenario modelling, highlighting some limitations that warrant further investigation.

5.1 ENCOUNTERS AS A PREDICTOR FOR COLLISIONS

The results are mixed in the predictive power of encounters and collision risk, with Pearson correlations between 0.61 and 0.18. This result is perhaps not surprising given the inherent complexity in predicting accidents in a highly complex system, with numerous inter-relating contributory factors (Du et al. 2020). In addition, the high contribution of human factors to historical collisions is well known, which cannot be directly observed from AIS data. Other external factors which might influence this relationship are weather conditions, waterway geometry, hydrodynamic effects and the presence of risk controls such as pilotage (Mazaheri et al. 2014; Bye and Aalberg, 2018; Olba et al. 2019). Further work is proposed to better understand whether the inclusion of these variables when modelling collision risk achieves better results.

In addition, the implementation of this approach treats encounters as pairwise interactions between vessels, which may result in a collision. As a result, it fails to capture complex multi-vessel encounter situations, which might have a greater likelihood of resulting in a collision and can only be evaluated using alternative methods (Zhang et al. 2020).

The quality of data should also be considered. The MarineCadastre does not have complete AIS coverage of all areas of the US coastline, therefore where this coverage is worst, there would be a relatively low number of encounters given reported collisions. Similarly, the further offshore, AIS reception degrades and a similar result would be expected. This might also apply for certain vessel types, with fishing and recreational having lower AIS carriage rates than commercial shipping. In addition, the relative under-reporting of incidents is well known (Qu et al 2012; Hassel et al. 2011), which makes establishing correlations difficult, particularly for non-commercial vessels.

Whilst the correlations of the models are varied, its implementation in characterising where collisions are more likely shows that collision risk can be characterised across regions. This supports decision makers in determining where risk is highest and implementing targeted risk controls to reduce that risk. Yet, our results also show that vessel traffic exposure can achieve almost equivalent results in some vessel types. Some relationship between vessel traffic activity and the number of activities is to be expected, yet the strength of this relationship suggests that simple measures of vessel movements can be used to characterise collision risk. This approach has numerous advantages as it requires less technical input, less computation and therefore has greater utility for navigation authorities.

Theoretically, we consider that the principal advantage of domain modelling over other models (such as geometric route intersections) is the inclusion of the time element. Therefore, a domain model would better represent a system where there is significant variability in spatiotemporal activity, such as where a day/night or tidal aspect is significant. In addition, in our future case scenario where the underlying vessel traffic volume remains unchanged, domain analysis is able to show not only a quantitative increase in the number of encounters, but where those additional encounters occur. This provides much more targeted information than traffic volume or incident rates could provide.

5.2 SUITABILITY OF DOMAIN MODEL

As a result of the multitude of domain shapes available for implementation by an analyst seeking to predict collision risk (Fiskin et al. 2020), it is likely that domain shape has some influence on the results (Goerlandt and Kujala, 2014). Given that there is no agreed ship domain concept, some limitations are inevitable on our adoption of Wang (2010), and these are briefly commented upon below.

Firstly, as has have been summarised by Zhang et al. (2016) there are certain limitations with oval domain shapes more generally. Ovals are symmetrical and therefore suggests that the influence of the COLREGs, which take into account the direction that meeting vessels should cross, is asymmetrical, and previous empirical analysis of ship domains has shown this (Hansen, 2013). Where the domain extends behind a vessel, particularly in Fujii and Tanaka (1971) where the forward and aft domains are equal, this suggests that crossing astern of a vessel is equal to crossing ahead of a vessel. In some cases, including in our implementation, it is possible for one vessel to encounter another, but not vice versa. For example, a large ship with a large domain would encounter a passing fishing vessel, but due to their much smaller domain, they would not encounter the ship. In this situation, we record an encounter, but some implementations may suggest that the risk of collision lies with the ship only and not the fishing vessel, when clearly one would necessarily collide with the other.

Some authors have presented non-binary domain models where the encounter is weighted by the significance of the encounter, with high speed and close encounters assigned more weight than slow speed and distant encounters. Whilst we have not included this aspect in our chosen implementation of Wang's (2010) model, we do not believe this would significantly alter the spatial distribution of the results. We expect instead that congested waterways where ships come close together would have both more encounters and those encounters would naturally be closer together on average. Conversely open waters enable masters to increase their passing distance, resulting in less encounters and at a greater distance on average. An interesting point is raised here as to whether such an approach adds any value to strategic use of domain analysis in collision risk assessment, given that the size and shapes of the domains already vary based on the size and speed of the vessel. The authors are not aware of any work that concludes that non-binary domain encounters are better predictors to the frequency of collisions than binary domain encounters, but this could be investigated as part of future work.

The formulas provided by many authors fail to account for many environmental and circumstantial details which might alter the natural separation between vessels. Reflective of this are the great number of studies that have focused on applications within a single waterway, in order to demonstrate their proposed domain shape. For example, Rawson et al. (2014) propose a shape fitted to navigation on the River Thames, which is necessarily different to ones proposed for open sea (Fujii and Tanaka, 1971), port approaches (Bakdi et al. 2019; Fang et al. 2019) or TSS (Zhang and Meng, 2019). The influence of waterway characteristics is therefore a major determining factor of navigation behaviour and therefore domain shape (Szlapczynski and Szlapczynska, 2017). Some have argued that a universal ship domain might be achievable (Pietrzykowski and Uriasz, 2009), applicable to all waterways, but this may not be possible or even desirable if they are applied for specific purposes.

This effect may be important, in constrained waterways where vessels naturally navigate closer together. Should the size of the domain reduce to account for the routine proximity at which vessels operate to differentiate normal encounters to abnormally close encounters. Similarly, it could be argued that as vessels navigate close together, bridge teams compensate by increased alertness, maybe even navigating with pilots, and therefore the awareness and reaction capability is far quicker, supporting smaller domains than at open sea. By contrast, more collisions occur in ports and constrained waters, supporting the association between collisions and encounters. Further work is planned in order to investigate the impact of waterway characteristics on empirically derived domain shapes.

5.3 DETERMINING INCREASED COLLISION RISK IN FUTURE SCENARIOS

As described in Section 4.3, by re-routeing vessel traffic around a proposed development, we can predict how the number of encounters might increase. Whilst we have raised questions on the relationship between number of encounters and historical collisions; in this situation, given that the traffic volume between the base and future case situations is the same, domain analysis has the advantage of being able to characterise how risk changes, something that traffic volume alone could not achieve.

An outstanding issue is the degree to which the increase in encounters is reflective of an increase in collision risk, rather than artificial encounters as a result of the modelling methodology. For example, if we take the number of encounters as E, then the increase in encounters takes the form:

$$\Delta E = E_{Modelled} - E_{Baseline} \tag{7}$$

In the baseline case, vessels are avoiding one another and attempting to maintain domain separation. Once traffic is re-routed, the ships may be assigned new paths which bring them into conflicts with other navigating vessels. Whilst this is useful, as it indicates that two vessels would now meet when they would not previously, there is an unknown likelihood that one of those vessels would have taken action to increase the separation distance, thereby avoiding an encounter. Some factor P_A, the probability of avoidance, is necessary to adjust the resulting increases in encounter situations.

It should be noted that this is distinct from the P_C or causation probability used in other risk assessment methods, which is the conditional probability of a collision given an encounter ($P_{Collision}|P_{Encounter}$), which might be in the region of 4.9×10^{-5} (Goerlandt and Kujala, 2014). Therefore, we should utilise the equation in the form to correct for this:

$$\text{Collisions} = (((E_{Modelled} - E_{Baseline}) * P_A) + E_{Baseline}) * \ P_c \tag{8}$$

This aspect is also relevant were some risk algorithm incorporated into the equation, such that collision risk is calculated per encounter. As proposed by some authors, inclusion of vessel speed, size and type can graduate the severity of the encounter situations. Yet, re-routed vessels following blind navigation may overlap with existing traffic, and have high speed, zero-distance encounters, that would generate artificially high collision risk scores.

P_A is an unknown factor, but some proposed methods for solving this are discussed below. Firstly, this could be estimated using expert judgement based on the perceived probability of avoiding another approaching vessel. Secondly, by comparing the rate of encounters to transits in other waterways which have similar characteristics as that of the diverted scenario, may provide some indication of the chance of two vessels encountering in compressed waterways. Finally, the baseline scenario could itself be modelled, by generating the baseline traffic along a route, the difference between the actual baseline and the modelled baseline reflects new encounters generated due to the modelling alone.

A secondary point when applying this methodology is the size of the study area. In our case study, we utilised only a small section of the Strait of Juan de Fuca, and therefore the impact on collision risk was localised. As a result, the relative increase in risk in this small section is significant. If the results are replicated for the whole Strait, then the relative increase would be much less. This would impact upon stakeholder reception of the analysis results, presenting high or low relative increases might influence their decision making of what is acceptable. This therefore requires some degree of presentation of risk in both absolute and relative terms, which domain analysis would facilitate.

6 CONCLUSIONS

This paper has investigated the practical applications of domain analysis in strategic risk assessments and critiqued their utility at predicting present day and future case collision risk. Whilst other methods of collision risk assessment exist, the widespread adoption of domain analysis within both academia and industry warrants a detailed evaluation to ensure it is fit for purpose. In testing this, our work has achieved the following key findings. Firstly, the statistical relationship between encounters and collisions is not strong, particularly for certain vessel types. This suggests that, whilst this approach may be useful in determining where collision risk is highest, the inclusion of other variables such as environmental conditions and local practice may be necessary to characterise the risk more accurately. This has been attempted by some authors, but the improvement in the predictive accuracy of these approaches has not been demonstrated. Secondly, as with other spatial models where data is aggregated, the statistical relationship is impacted by the MAUP, whereby the scale of assessment impacts the statistical relationships derived. Such an effect must therefore be considered by authors when employing this approach. Thirdly, we demonstrate that using a measure of vessel activity has predictive accuracy equivalent to frequency of encounters in some cases as a predictor of collision frequency. This suggests that simple methods, such as volume of traffic, may be as accurate as more complex methods for use in maritime risk assessment, with significantly less computational requirements.

These findings aside, in addition we have demonstrated the applicability of domain analysis to quantify the impact of a development or obstacle on collision risk. This supports evidence-based decision making on the magnitude of risks, and the effectiveness of proposed mitigation, during the permitting application and environmental studies of projects such as wind farms. Yet, several key challenges with this approach are highlighted that require additional research. Most notably the action of diverting the vessel traffic to avoid an obstruction generates a significant number of encounters which unrealistically do not avoid one another. Resolving this limitation is essential before domain analysis's strong capability to predict changes in risk can be realised.

ACKNOWLEDGEMENTS

This work is partly funded by the University of Southampton's Marine and Maritime Institute (SMMI) and the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement number: 723526: SEDNA).

REFERENCES

Anatec, 2019. Thanet Offshore Wind Farm: Collision Assessment of Proposed Extension. https://infrastructure.planninginspectorate.gov.uk/wp-content/ipc/uploads/projects/EN010084/EN010084-001987-Vattenfall%20Wind%20Power%20Limited%20-%20D6_Appendix42_TEOW_CRM_RevA.pdf (accessed 18 February 2020).

Ari, I., Aksakalli, V., Aydogdu, V., Kum, S., 2013. Optimal Ship Navigation with Safety Distance and Realistic Turn Constraints. European Journal of Operational Research 229(3), 707-717. https://doi.org/10.1016/j.ejor.2013.03.022.

Aven, T., Heide, B. 2009. Reliability and validity of risk analysis. Reliability Engineering and System Safety 94, 1862-1868. https://doi.org/10.1016/j.ress.2009.06.003.

Bakdi, A., Glad, I., Vanem, E., Engelhardtsen, O., 2019. AIS-based Multiple Vessel Collision and Grounding Risk Identification based on Adaptive Safety Domain. Journal of Marine Science and Engineering 8, 5. https://doi.org/10.3390/jmse8010005.

Barnes, R., 2016. dggridR: Discrete Global Grid Systems for R. https://github.com/r-barnes/dggridR (accessed 24 May 2019).

Bye, R., Aalberg, A., 2018. Maritime navigation accidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports. Reliability Engineering and System Safety, 176, 174-186. https://doi.org/10.1016/j.ress.2018.03.033.

Bye, R., Almklov, P., 2019. Normalization of maritime accident data using AIS. Marine Policy 109, 103675. https://doi.org/10.1016/j.marpol.2019.103675.

Chen, P., Huang, Y., Mou, J., van Gelder, P., 2019. Probabilistic risk analysis for ship-ship collision: State of the art. Safety Science 117, 108-122. https://doi.org/10.1016/j.ssci.2019.04.014.

Correndo, G., 2019. DGGRIDPY Git-hub page. https://github.com/correndo/dggridpy (accessed 22 May 2020).

Du, L., Goerlandt, F., Kujala, P., 2020. Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from AIS data. Reliability Engineering and System Safety 200, 106933. https://doi.org/10.1016/j.ress.2020.106933.

EMSA, 2019. Annual Overview of Marine Casualties and Incidents 2019. http://www.emsa.europa.eu/emsa-homepage/2-news-a-press-centre/news/3734-annual-overview-of-marine-casualties-and-incidents-2019.html. (accessed 18 May 2020).

Fang, Z., Yu, H., Ke, R., Shaw, S., Peng, G., 2019. Automatic Identification System-Based Approach for Assessing the Near-Miss Collision Risk Dynamics of Ships in Ports. IEEE Transactions on Intelligent Transportation Systems 20(2), 534-543. https://doi.org/10.1109/TITS.2018.2816122.

Feng, Z., Yang, H., Li, X., Li, Y., Liu, Z., Liu, R., 2019. Real-time Vessel Trajectory Data-Based Collision Risk Assessment in Crowded Inland Waterways. 4th IEEE International Conference on Big Data Analytics, 128-134. https://doi.org/10.1109/ICBDA.2019.8712843.

Fiskin, R., Nasiboglu, E., Yardimci, M., 2020. A knowledge-based framework for two-dimensional (2D) asymmetrical polygonal ship domain. Ocean Engineering, 202. https://doi.org/10.1016/j.oceaneng.2020.107187.

Fournier, M., Hilliard, C., Rezaee, S., Pelot, R., 2018. Past, present and future of the satellite-based automatic identification system: areas of applications (2004-2016). WMU Journal of Maritime Affairs 17, 311-345. https://doi.org/10.1007/s13437-018-0151-6.

Fujii, Y., Tanaka, R., 1971. Traffic Capacity. Journal of Navigation 24, 543-552. https://doi.org/10.1017/S0373463300022384.

Goerlandt, F., Kujala, P., 2014. On the reliability and validity of ship-ship collision risk analysis in light of different perspectives on risk. Safety Science 62, 348-365. https://doi.org/10.1016/j.ssci.2013.09.010.

Goodwin, E. 1975. A statistical study of ship domains. Journal of Navigation 28, 328-344. https://doi.org/10.1017/S0373463300041230.

Hanninen, M., 2014. Bayesian Networks for Maritime Traffic Accident Prevention: Benefits and Challenges. Accident Analysis and Prevention 73, 305-312. https://doi.org/10.1016/j.aap.2014.09.017.

Hansen, M.G., Jensen, T.K., Lehn-Schiøler, T., Melchild, K., Rasmussen, F.M., Ennemark, F., 2013. Empirical Ship Domain based on AIS Data. Journal of Navigation 6, 931-940. https://doi.org/10.1017/S0373463313000489.

Harati-Mokhtari, A., Brooks, P., Wall, A., and Wang, J. (2007). Automatic Identification System (AIS): Data Reliability and Human Error Implications. Journal of Navigation 60, 373-389. https://doi.org/10.1017/S0373463307004298.

Hassel, M., Asbjornslett, B., Hole, L., 2011. Underreporting of maritime accidents to vessel accident databases. Accident Analysis and Prevention 43, 2053-2063. https://doi.org/10.1016/j.aap.2011.05.027.

Heinrich, H. 1931. Industrial Accident Prevention. A Scientific Approach. McGraw-Hill Book Company, Inc. New York.

IALA, 2002. IALA Guidelines on the Universal Automatic Identification System (AIS). Volume 1, Part II – Technical Issues. Edition 1.1.

Im, N., Luong, T., 2019. Potential risk ship domain as a danger criterion for real-time ship collision risk evaluation. Ocean Engineering 194, 106610. https://doi.org/10.1016/j.oceaneng.2019.106610.

Kulkarni, K., Goerlandt, F., Li, J., Banda, O., Kujala, P., 2020. Preventing shipping accidents: Past, present, and future of waterway risk management with Baltic Sea focus. Safety Science 129, 104798. https://doi.org/10.1016/j.ssci.2020.104798.

Li, S., Meng, Q., Qu, X., 2012. An Overview of Maritime Waterway Quantitative Risk Assessment Models. Risk Analysis 32(3), 496-512. https://doi.org/10.1111/j.1539-6924.2011.01697.x.

MAIB, 2014. Report on the investigation of the collision between Paula C and Darya Gayatri in the South-west lane of the Dover Traffic Separation Scheme on 11 December 2013. Report No 25/2014.

Marico Marine, 2018. Thanet Extension Offshore Wind Farm: Navigation Risk Assessment. https://infrastructure.planninginspectorate.gov.uk/wp-content/ipc/uploads/projects/EN010084/EN010084-000606-6.2.10_TEOW_S&N.pdf (accessed 18 February 2020).

MarineCadastre, 2020. Vessel Traffic Data. https://marinecadastre.gov/ais/ (accessed 15 May 2020).

Mazaheri, A., Ylitalo, J., 2010. Comments on Geometrical Modelling of Ship Grounding. 5th Conference on Collision and Grounding of Ships, Espoo, Finland, 14-16 June 2010.

Mazaheri, A., Montewka, J., Kujala, P., 2013. Modelling the risk of ship grounding – a literature review from a risk management perspective. WMU Journal of Maritime Affairs 13(2), 269-297. https://doi.org/10.1007/s13437-013-0056-3.

Mazaheri, A., Montewka, J., Kotilainen, P., Sormunen, O.E., Kujala, P., 2014. Assessing Grounding Frequency using Ship Traffic and Waterway Complexity. Journal of Navigation 68(1), 89-106. https://doi.org/10.1017/S0373463314000502.

Olba, X., Daamen, W., Vellinga, T., Hoogendoorn, S., 2019. Risk Assessment Methodology for Vessel Traffic in Ports by Defining the Nautical Port Risk Index. Journal of Marine Science and Engineering 8, 10. https://doi.org/10.3390/jmse8010010.

Pedersen, P. T., 1995. Collision and Grounding Mechanics. Proceedings of WEMT 95, Copenhagen, Denmark, The Danish Society of Naval Architecture and Marine Engineering.

Pietrzykowski, Z., Uriasz, J., 2009. The Ship Domain – A Criterion of Navigational Safety Assessment in an Open Sea Area. Journal of Navigation 62, 93-108. https://doi.org/10.1017/S0373463308005018.

Qu, X., Meng, Q., Li, S., 2011. Ship collision risk assessment for the Singapore Strait. Accident Analysis and Prevention 43(6), 2030-2036. https://doi.org/10.1016/j.aap.2011.05.022.

Qu, X., Meng, Q., Li, S., 2012. Analyses and Implications of Accidents in Singapore Strait. Journal of Transportation Research Board 2273, 106-111. https://doi.org/10.3141/2273-13.

Rawson, A., Rogers, E., Foster, D., Phillips, D., 2014. Practical Application of Domain Analysis: Port of London Case Study. Journal of Navigation 67, 193-209. https://doi.org/10.1017/S0373463313000684.

Rawson, A., Sabeur, Z., Correndo, G., 2019. Spatial challenges of maritime risk analysis using big data. Soares, C. Guedes (ed.) In Proceedings of the 8th International Conference on Collision and Grounding of Ships and Offshore Structures (ICCGS 2019). vol. 4, CRC Press / Balkema, 275-283.

Rong, H., Teixeira, A., Soares, C., 2019. Risk of ship near collision scenarios off the coast of Portugal. Proceedings of the 29th European Safety and Reliability Conference, Germany, 3660-3666.

Sahr, K.M., White, A.J., 1998. Discrete Global Grid Systems. Weisberg, S. (ed.) In Computing Science and Statistics, 30, 269-278.

Sahr, K.M., White, D., Kimerling, A.J., 2003. Geodesic Discrete Global Grid Systems. Cartography and Geographic Information Science 30(2), 121-134. https://doi.org/10.1559/152304003100011090.

SciPy. 2020. SciPy documentation: scipy.stats.pearsonr. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html (accessed 07 August 2020).

Szlapczynski, R., Szlapczynska, J., 2017. Review of ship safety domains: Models and applications. Ocean Engineering 145, 277-289. https://doi.org/10.1016/j.oceaneng.2017.09.020.

Tam, C., Bucknall, R., Greig, A., 2009. Review of Collision Avoidance and Path Planning Methods for Ships in Close Range Encounters. Journal of Navigation 62, 455-476. https://doi.org/10.1017/S0373463308005134.

USCG, 2020. Marine Casualty and Pollution Data for Researchers. https://www.dco.uscg.mil/Our-Organization/Assistant-Commandant-for-Prevention-Policy-CG-5P/Inspections-Compliance-CG-5PC-/Office-of-Investigations-Casualty-Analysis/Marine-Casualty-and-Pollution-Data-for-Researchers/ (accessed 29 May 2020).

Van Dorp, J.R., Harrald, J.R., Marrick, J.R.W., and Grabowski, M. 2008, VTRA: Technical Appendix D: Expert Judgement Elicitation. https://www2.seas.gwu.edu/~dorpjr/VTRA/FINAL%20REPORT/083108/ VTRA%20REPORT%20-%20Appendix%20D%20083108.pdf (accessed 19 April 2019).

Van Dorp, J., Merrick, J., 2014. VTRA 2010 Final Report. George Washington University.

Wang, N., Meng, X., Xu, Q., Wang, Z., 2009. A Unified Analytical Framework for Ship Domains. Journal of Navigation 62, 643-655. https://doi.org/10.1017/S0373463309990178.

Wang, N., 2010. An Intelligent Spatial Collision Risk Based on the Quaternion Ship Domain. Journal of Navigation 63, 733-749. https://doi.org/10.1017/S0373463310000202.

Xu, Q., Wang, N. 2014. A Survey on Ship Collision Risk Evaluation. Traffic Management Review 26, 475-486. https://doi.org/10.7307/ptt.v26i6.1386.

Yang, D., Wu, L., Wang, S., Jia, H., Li, K., 2019. How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications. Transport Reviews, 39, 755-773. https://doi.org/10.1080/01441647.2019.1649315.

Zhang, W., Goerlandt, F., Kujala, P., Wang, Y., 2016. An advanced method for detecting possible near miss ship collisions from AIS data. Ocean Engineering 124, 141-156. https://doi.org/10.1016/j.oceaneng.2016.07.059.

Zhang, L., Meng, Q., 2019. Probabilistic ship domain with applications to ship collision risk assessment. Ocean Engineering 186, 106130. https://doi.org/10.1016/j.oceaneng.2019.106130.

Zhang, W., Feng, X., Goerlandt, F., Liu, Q., 2020. Towards a Convolutional Neural Network model for classifying regional ship collision risk levels for waterway risk analysis. Reliability Engineering and System Safety 204. https://doi.org/10.1016/j.ress.2020.107127.

A CRITIQUE OF THE USE OF DOMAIN ANALYSIS FOR SPATIAL COLLISION RISK ASSESSMENT

Andrew Rawson[a*] and Mario Brito[b]

[a] Electronics and Computer Science, University of Southampton, SO17 1BJ, UK (A.Rawson@soton.ac.uk) – Corresponding Author

[b] Decision Analytics and Risk, Southampton Business School, University of Southampton, SO17 1BJ, UK (M.P.Brito@soton.ac.uk)

ABSTRACT

Predicting the likelihood of maritime ~~incidents~~ accidents is hindered by the relative sparsity of collisions on which to develop risk models. Therefore, significant research has investigated the capability of non-accident situations, near misses and encounters between vessels as a surrogate indicator of collision risk. Whilst many studies have developed ship domain concepts, few have considered the practical considerations of implementing this method to ~~utilised ship domains as a method of characterising~~ characterise navigational risk between waterways and scenarios. In order to address this, within this paper we implement and evaluate the capability and validity of domain analysis to characterise and predict the likelihood of ship collisions. Our results suggest that the strength of the relationship between collisions and encounters is varied both between vessel types and the spatial scale of assessment. In addition, we demonstrate some key practical considerations in utilising domain analysis to predict~~ing~~ the change in collision risk, through a hypothetical wind farm. The outcomes of this study provide research direction for practical applications of domain analysis on collision risk assessments.

KEYWORDS

Ship domain, Collision Risk Assessment, Automatic Identification System

1 INTRODUCTION

Tools to assess and predict the likelihood or consequences of vessel accidents have been key focus of many authors. Comprehensive reviews of these methods have been provided by many (Li et al. 2012, Xu and Wang 2014, Chen et al. 2019, Kulkarni et al. 2020) and include ~~various~~ statistical derivation of incident rates (Bye and Almklov, 2019), analytical models (Pedersen, 1995; Mazaheri and Ylitalo, 2010; Li et al. 2012), or the use of Bayesian Networks (Hanninen, 2014). Assessing the accuracy of these~~These~~ models ~~are~~ is challenged by the ~~relative~~ infrequency at which accidents occur relative to the volume of traffic within an area. For example, even in busy areas such as the Dover Straits, the annual collision rate is approximately 1.2 incidents per year (MAIB, 2014). Given that collisions are a leading cause of ship casualties, accounting for 26% of incidents (EMSA, 2019) and potentially resulting in loss of life and pollution, alternative methods have been proposed to assess collision risk.

To overcome this, many authors have proposed the use of non-critical accident situations, near misses or encounters as a proxy measure for vessel risk (Du et al., 2020). As relates to the risk of collision, where two navigating vessels come into contact with one another, a widely adopted approach is that of the ship domain. This concept represents the surrounding effective waters which a navigator wishes to keep clear of other ships or fixed objects (Goodwin, 1975). Where the domain of a vessel is ~~breached~~violated, a threat to navigational safety has occurred (Pietrzykowski and Uriasz, 2009). These domains can range from simple circular buffers to more complex, segmented and dynamic shapes (Fiskin et al. 2020). The size and shape can change depending on several factors including the physical characteristics of the vessels, the encounter situation (head-on, crossing or overtaking as defined by the Collision Regulations), manoeuvrability, the human element, metocean conditions and fairway characteristics. Fiskin et al. (2020) provide a particularly good comparison of the different factors taken into consideration in various domain models. Whilst the terminologies used in previous work vary (Du et al., 2020), the underlying premise is that for a collision to occur, two vessels must meet, and by determining where two vessels meet more often, the likelihood of a collision increases.

A ship domain can be approached in one of two ways. Firstly, as a tactical tool for collision avoidance, whereby a minimum domain is desired and therefore an intelligent navigation system might seek to maintain such passing distance when interacting with other vessels (see for example Tam et al. 2009; Im and Luong, 2019). Secondly, as a strategic tool for collision risk assessments, whereby the frequency of domain interactions is used as a proxy measure for collision risk. This enables quantitative analysis of collision risk, supporting a proactive approach to risk management.

This latter concept has parallels to the Heinrich (1931) accident pyramid, similar accident chains have varying levels of severity. By identifying and preventing lesser incidents, such as unsafe acts, more serious incidents such as fatalities could be avoided. He postulated that for each 300 unsafe acts, 29 minor injuries and one major injury or fatality would occur. Within the concept of vessel collisions, the same causal events which lead to domain encounters, would lead to collisions, and therefore determining the frequency of encounters can inform collision risk.

Within the academic literature, a plethora of domain shapes and sizes has been proposed (Wang et al. 2009; Pietrzykowski and Uriasz, 2009; Szlapczynski and ~~Szlapczynski~~Szlapczynska, 2017; Bakdi et al. 2019; Fiskin et al. 2020). Yet, there are relatively few examples of the application of domain analysis to characterise collision risk across waterways (Szlapczynski and Szlapczynska~~i~~, 2017), or more generally some have criticised the lack of practical applications of maritime risk models (Kulkarni et al. 2020). It is seemingly an accepted logic that where more domain encounters occur, more collisions should be expected, yet few studies have tested this link (Zhang et al. 2016; Du et al. 2020). Furthermore, several authors have challenged the predictive capability of these models (Goerlandt and Kujala, 2014), having tested the reliability and validity of different approaches. Indeed, the significant diversity of methods proposed can be seen as evidence that there is no single ship domain that can be applied to all circumstances, and a diversity of results should therefore be expected. In addition, most studies have been conducted limited to small areas with homogenous characteristics, such as individual ports (Fang et al. 2019) or isolated waterways (Fujii and Tanaka, 1971).

These issues are not purely academic, with ship domain theory commonly applied to maritime risk assessments in applied settings. Within this context, by comparing the base case number of encounters with an altered future case, the relative difference represents a change in the risk of collision. Such an outcome can then be used to inform consenting decisions by regulators or the cost benefit of an expensive ~~costly~~ risk control. For example, ~~S~~ship domains have been used to assess the navigational risk implications of infrastructure such as offshore wind farms (Marico Marine, 2018) or oil and gas terminals (van Dorp et al., 2008; 2014). These navigation risk assessments (NRAs) are crucial in ensuring decision makers are able to identify, assess and evaluate maritime risk correctly. Given the outstanding research questions related to the use of ship domains for NRAs, there is a need to resolve these issues to avoid applications providing either erroneous or misleading results.

Within this paper we provide a critique of the use of domain analysis to assess navigational risk. ~~and provide several key contributions. Firstly, we~~We implement a simple domain at a national scale to demonstrate how collision risk could be characterised using this approach. From this, we then test the statistical relationship between the number of encounters and the historical number of collisions, to consider the validity of such an approach. Finally, we extend the analysis to consider how domain analysis could be implemented to predict the change in collision risk following a new development.

This work offers a number of key contributions. Firstly, we test the relationship between encounters and historical collisions using a statistical analysis at a national level. We demonstrate that there are few studies which have used domain analysis for this purpose, and fewer have evaluated the validity of this approach. Secondly, by utilising a national approach, using a significant volume of data, we are able to offer a more generalised approach than previous studies within specific waterways which have their own local conditions. Thirdly, we contribute a framework and case study through which this approach can be utilised to predict the change in collision risk following a development. Our work demonstrates some clear limitations which warrant further investigation.

This paper is structured as follows; Section 2 describes the key literature for domain analysis and collision risk. Section 3 describes the datasets, methodology and model implemented within this project. Section 4 describes the results of the analysis. A discussion is contained in Section 5 to ~~describe the results and~~consider the key findings of this assessment.

2 LITERATURE REVIEW

The statistical and geometric models widely utilising in the literature are static representations of flows of traffic and are therefore limited in their ability to represent the complexity of the maritime situation (Mazaheri et al. 2013). The increased capabilities of computing and increasing availability of vessel traffic data have allowed for a parallel stream of temporal-spatial modelling. Unlike geometric analysis, which characterises traffic as linear flows, the construction of simulations allows for the reflection of the dynamic nature of vessel behaviour. Through this approach, collision risk can be assessed by comparing the time and distance to closest point of approach or by constructing vessel domains and evaluating interactions and encounters between those domains. Amongst collision avoidance and near miss detection/trajectory processing, waterway risk assessment is one of the key applications of domain analysis (Szlapczynski and Szlapczynska, 2017). Whilst innovative new methods are continual proposed, such as the use of Convolutional Neural Networks and artificial intelligence to evaluate traffic situations (Zhang et al. 2020), ship domain analysis has achieved widespread adoption. Within this section we review the previous work on both domain modelling in collision risk assessment and the evaluation of maritime risk models.

~~2~~

Formatted: Normal

Formatted: Normal

2.1 DOMAIN MODELLING ~~AND~~ IN COLLISION RISK ASSESSMENT

~~The statistical and geometric models widely utilising in the literature are static representations of flows of traffic and are therefore limited in their ability to represent the complexity of the maritime situation (Mazaheri et al. 2013). The increased capabilities of computing and increasing availability of vessel traffic data have allowed for a parallel stream of temporal-spatial modelling. Unlike geometric analysis, which characterises traffic as linear flows, the construction of simulations allows for the reflection of the dynamic nature of vessel behaviour. Through this approach, collision risk can be assessed by comparing the time and distance to closest point of approach or by constructing vessel domains and evaluating interactions and encounters between those domains. Amongst collision avoidance and near miss detection/trajectory processing, waterway risk assessment is one of the key applications of domain analysis (Szlapczynski and Szlapczynskai, 2017). Whilst innovative new methods are continual proposed, such as the use of Convolutional Neural Networks and artificial intelligence to evaluate traffic situations (Zhang et al. 2020), ship domain analysis has achieved widespread adoption.~~

The general rationale of a domain analysis approach takes the form as Equation 1, where N_c is the number of encounters and P_c is the causation probability that an encounter becomes a collision (Goerlandt and Kujala, 2014).

$$f = N_c P_c \quad (1)$$

Therefore, in order to predict the collision risk, we must first implement a domain model that counts the number of encounters between vessels within an area. Secondly, we must determine a conditional probability that an encounter becomes a collision. This latter approach could be achieved statistically, by comparing the number of encounters per unit time with the historical incident record, or by utilising a Bayesian Network to include a variety of contributory factors. However, it is evident that a significant emphasis is placed on the number of encounters within this approach. Such approaches have been presented by several authors, but practical applications of domain analysis are generally limited within the academic literature (Rawson et al. 2014; Szlapczynski and Szlapczynska~~i~~, 2017).

Qu et al. (2011) present collision risk per leg in the Straits of Singapore by applying fuzzy domain analysis, in combination with measures of speed profiles. The results allow determination of which legs have higher collision risk profiles than others, but this is not compared to historical incident rates. Other approaches are taken by Fang et al. (2019) for Xiamen Port and Feng et al. (2019) in Yangtze Port. Whilst the methods vary, in both these cases, encounters are used to identify where collisions are more likely, spatially, across a port's waterway. These studies are isolated to single specific waterways, lacking generalised applicability to large areas.

Rong et al. (2019) used domain analysis to evaluate near collision scenarios off the Portuguese Coast using three months of vessel traffic. Their use of the Fujii and Tanaka (1971) method with a collision risk evaluation, identified 1,671 near collision scenarios. Whilst the results show the spatial variation in where collision risk is perceived to be highest, these results are not evaluated or validated against any other measures of collision risk. Zhang et al. (2016) conducted near miss modelling in the Gulf of Finland, using k-means clustering to determine types of encounters.

Rawson et al. (2014) presents a domain model applied for collision risk on the River Thames, developed through consultation with local skippers. In this case, not only is the spatial variation in encounters used to determine where collisions are more likely, the relationship between traffic volume and number of encounters is investigated to predict how collision risk might change with increased numbers of vessels navigating on the future. Unlike many of the studies described above, this sought to develop domain analysis from a descriptive to a predictive tool.

The motivation to predict changes in collision risk have led to the inclusion of domain analysis in a number of major applied studies. For example, in van Dorp et al. (2008; 2014), a simulation is developed of vessel traffic in the waters of Washington State, USA. Collision candidates are determined every minute of the simulation based on proximity, and a probabilistic causation probability assigned based on expert judgement elicitation. Upon this, the model results are compared having included new traffic associated with a proposed marine terminal, and then various risk control measures are implemented, and the benefits compared. Domain analysis in this context is a decision support tool to determine both the acceptability of the risks associated with the development but assess the cost benefit of risk controls such as escort towage.

Other practical implementations include the aforementioned NRAs for offshore wind farms in the UK, where the baseline and future case risks associated with the development can be compared (Marico Marine, 2018; Anatec, 2019). The relative change in encounters informs decision makers on the potential increase in risk associated with a development.

2.2 VERIFICATION OF MODEL RESULTS

An interesting aspect which is often omitted from studies on domain analysis, and is generally assumed, is how representative domain analysis results are for either characterising or predicting collision risk. Within their review, Du et al. (2020) note that there is comparatively little academic research focusing on the validity of non-accident critical event detection models. This is both a result of the relative sparsity of accidents upon which to validate the model, and the complexity of the accident scenario due to inter-relating contributory factors. ~~(Du et al. 2020).~~

Aven and Heide (2009) define model validity as the degree to which the risk analysis describes the specific concepts that one is attempting to describe, including amongst others, the degree to which the model risk scores reflect the underlying true risk. A significant contribution to framing this discussion is that of Goerlandt and Kujala (2014). They evaluated the reliability and validity of collision risk models in a Traffic Separation Scheme (TSS). Their results showed that by using different models, significantly different frequencies and spatial distributions of collision risk were obtained. This raises questions on the predictive capability of domain analysis for collision risk assessment. This can be further evidenced by the wide variety of models presented within the literature (Goerlandt and Kujala, 2014), with different shapes, functions regarding speed or encounter type and numerous other variables.

By way of example, the application for an offshore wind farm in the UK was assessed by two consultancies using two forms of domain analysis in order to measure the relative impact of the development on collision risk (Marico Marine, 2018; Anatec, 2019). One assessment presented an increase in collision risk of 50% from approximately once in six to approximately once in four years whilst the other reports a change from once in 47 years to once in 46 years, a 2.2% increase. These differences are significant, demonstrating that a great variability in results can be derived by alternative implementations of the same methodological approach.

One method to validate the results of an assessment might be to compare the model results with historical accident numbers, such that where most domain encounters occur, is where most accidents have occurred. Where accident data is unavailable, the validity of derived high encounter intensity regions can be compared against perceived high-risk regions by consulting with local practitioners. For example, in Rawson et al. (2014), the results were recognisable to local stakeholders, however, this does not necessarily support the underlying statistical validity. A further method would be to evaluate the method against some other method, for example, Zhang et al. (2016) compare their encounter results against previous studies in the same area but using different methods.

Within this paper we seek to test this relationship and provide some research direction on the application of domain analysis for collision risk assessments. We ask two key questions; firstly, what is the validity in the implementation of domain analysis to characterise maritime risk spatially? We achieve this by analysing domain encounters across the continental United States during June 2018. Secondly, what are the research challenges in utilising this approach to predict the change in risk following future changes in traffic activities? In order to test this second aspect, we utilise a hypothetical case study of a new development in the Strait of Juan de Fuca, Washington State.

3 METHODOLOGY

~~3~~ Within this section, the methodological approach is described, including the underlying datasets, utilised domain function and statistical method to determine the relationship between encounters and collisions. In addition, we present a framework for utilising encounters to assess the impact of future scenarios.

Formatted: Normal

3.1 DATASETS

3.1.1 Vessel Traffic Data

Vessel traffic data from the Automatic Identification System (AIS) was used to model the navigation of vessels within the study area. AIS is a transponder system required to be fitted to all commercial vessels, under SOLAS Chapter V, and voluntarily carried by smaller craft, that sends and receives information about the dynamic movement (~~location~~position, speed, course and others) and static identification (type, size, name and others) of navigating vessels (IALA, 2002). Whilst the system was principally developed for improving maritime safety, AIS data enables high resolution analysis across a wide range of applications (Fournier et al. 2018; Yang et al. 2019).

The Marine Cadastre (2020) publish AIS data collected the US Coast Guard's national network of AIS receivers. Data was extracted from the 1st June to 30th June 2018, containing 32GB of AIS data in csv format. The dataset was processed to filter to the study area and linear interpolation used to standardise the dataset to one position per vessel per minute, provided the time between sequential positions of the same vessel did not exceed ten minutes, whereby it is assumed that tracking of the vessel has been lost and the interpolated positions have insufficient accuracy for use in collision risk assessment.

As AIS often contains missing or erroneous descriptions of vessels (Harati-Mokhtari et al. 2007), additional datasets were connected to that vessel containing summaries such as vessel type and length. Missing values were either filled in or if no vessel type information was found, the vessel was omitted from the final dataset. The resulting vessel traffic dataset contained approximately 120 million vessel positions and a density plot of the collected data is shown in Figure 1.

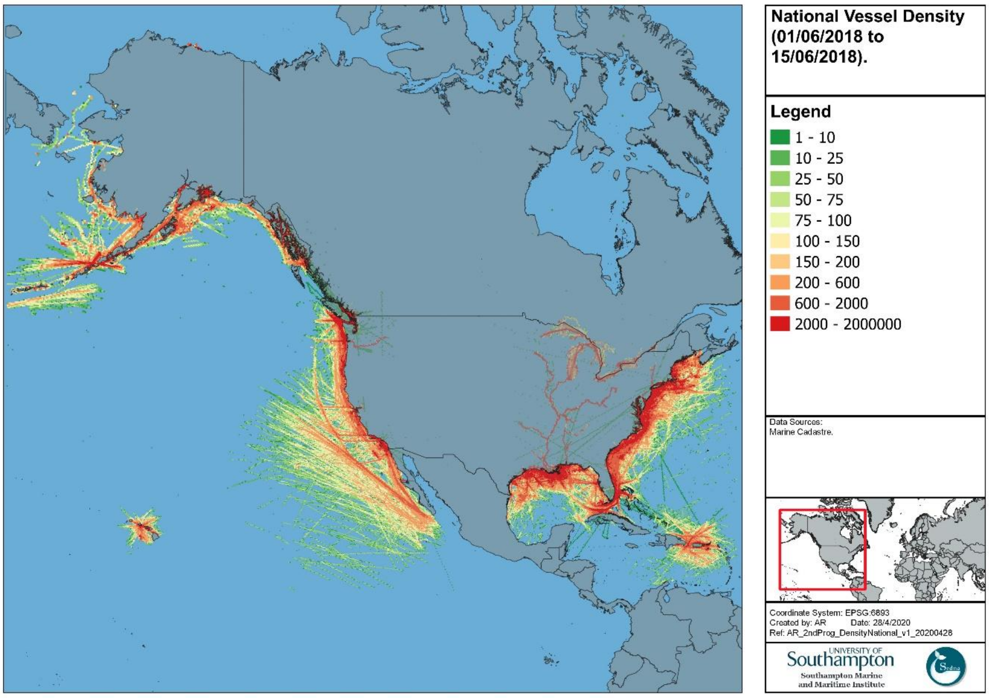


Figure 1: Vessel traffic density.

3.1.2 Incident Data

Under the Code of Federal Regulations 46 CFR 4.03 / 4.05, any marine casualty or accident occurring within the United States navigable waters, including grounding, collision, allision or flooding, shall be reported to the Coast Guard. A database of these incidents from 2002 to July 2015 is available specifically for use by researchers (USCG, 2020).

Two key files are relevant, the MisleVslEvents contains 132,717 entries and 20 fields for incidents involving vessels and the MisleVessel contains the details for 1.35 million vessels with 66 fields. Both datasets contain a vessel_id field which enables joining ship attribute data and incident data. The dataset was filtered to collisions only, often leaving two records per incident (signifying two or more vessels involved in a collision). In total, 5,165 instances of vessels involved in a collision across 2,524 unique collisions were used, with the locations of collisions shown in Figure 2. Most of the collisions are located at the locations of some of the country's major ports and along the Mississippi River.

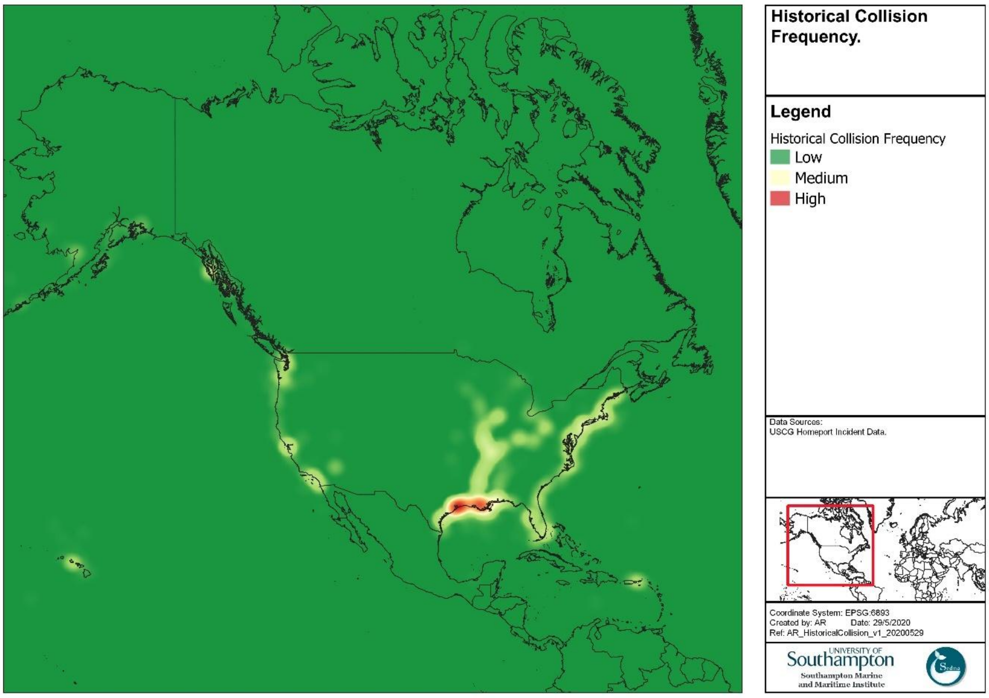


Figure 2: Heatmap of historical collisions in US.

3.2 DOMAIN MODEL

There are a significant number of different domain models from which to choose for a given study. Given the advantages discussed below, we choose to implement the domain model proposed by Wang (2010), with some modifications, within this section. Firstly, based on the reviews of Szlapczynski and Szlapczynska (2017), and Fiskin et al. (2020), we note that dynamic domains which change based on conditions, most commonly vessel length and speed, are popular amongst different authors and therefore our chosen domain model should reflect this. Furthermore, we do not wish to include many other factors as this increases the computational complexity, particularly given the significant scale of our assessment. Secondly, we wish to utilise a domain model which has been widely discussed in the literature and there are other studies utilising this approach, such as Qu et al.

(2012), and Goerlandt and Kujala (2014). Thirdly, the model must be well documented and therefore reproducible. ~~Whilst there are a wide variety from which to choose, the literature review has recommended that the domain should be dynamic, taking into account speed and vessel size. The domain proposed by Wang (2010) meets these traits and therefore has been implemented in this section.~~

Several authors, including Wang (2010), have proposed non-binary collision models whereby the significance of encounters is graduated. By contrast, binary domain models consider that two vessels either encounter or do not encounter, given the rules of the domain, and therefore one encounter is equal to another. One method to achieve a non-binary domain is to utilise fuzzy boundaries, whereby vessels that are outside of the fuzzy boundary are not at risk, but as the vessel is closer inside the fuzzy boundary the risk increases (Pietrzykowski and Uriasz, 2009). Another approach is to apply what is often referred to as a Vessel Conflict Ranking Operator (VCRO) which grades the severity of different types of encounter, for example, Zhang et al. (2016). The use of non-binary domains is useful in situations of collision avoidance, such that as two vessels come closer together, the risk of a collision increases and therefore the need for a response is heightened. However, in terms of strategic collision assessments, few studies have suggested that where more significant encounters occur the likelihood of a collision increases. Indeed, it could be argued that a properly calibrated dynamic domain should expand or contract ~~to ensure any encounter is of equal significance~~such that the domain boundary is of equivalent risk between different situations. Therefore, we have chosen not to implement a fuzzy or non-binary ship domain, in order to maintain simplicity and efficient computation in this comparative model, however we discuss the implications of this decision in Section 5~~.1~~.

In Wang (2010), the author proposes an Intelligent Spatial Collision Risk Model based on the Quaternion ship domain. The domain is an ellipse shape with four directional radius R_{fore}, R_{aft}, R_{port}, R_{starb} (see Figure 3 and Figure 4). The size of these radius are given by the formulas:

$$
\begin{cases}
R_{fore} = \left(1 + 1.34\sqrt{k_{AD}^2 + (k_{DT})/2)^2}\right)L \\
R_{aft} = \left(1 + 0.67\sqrt{k_{AD}^2 + (k_{DT})/2)^2}\right)L \\
R_{starb} = (0.2 + k_{DT})L \\
R_{port} = (0.2 + 0.75k_{DT})L
\end{cases}
\quad (2)
$$

299 Where L is the ship length and k_{AD} and k_{DT} represent the advance and tactical diameter of the vessel 300 scaled by some factor k that represents risk appetite (defaulted to 1), and where V_{own} is the ship 301 speed in knots, as given by:

$$
\begin{cases}
k_{AD} = \dfrac{A_D}{L} = 10^{0.3591 \lg V_{own} + 0.0952} \\
k_{DT} = \dfrac{D_T}{L} = 10^{0.5441 \lg V_{own} - 0.0795}
\end{cases}
\quad (3)
$$

302 To calculate whether a target vessel falls with own vessel domain, a pairwise distance calculation is 303 conducted at every timestamp, with distance between vessels i and j with coordinates X and Y (in 304 UTM metres) is derived by:

$$
D = \sqrt{(X_j - X_i)^2 + \ (Y_j - Y_i)^2} \quad (4)
$$

305 The identify function as to whether a target vessel falls within the domain can then be described as 306 (Zhang et al. 2016):

$$
l_\alpha = \left(\frac{1 + tan^2\alpha}{\frac{1}{S^2} + \frac{tan^2\alpha}{R^2}} \right) \quad (5)
$$

307 Where l_α is the distance to domain boundary at angle α, S is the R_{port} and R_{starb} radius functions and R 308 is the R_{fore}, R_{aft} functions. Therefore, a target vessel at relative bearing α from own vessel is within 309 the domain if the distance D is greater than l_α.

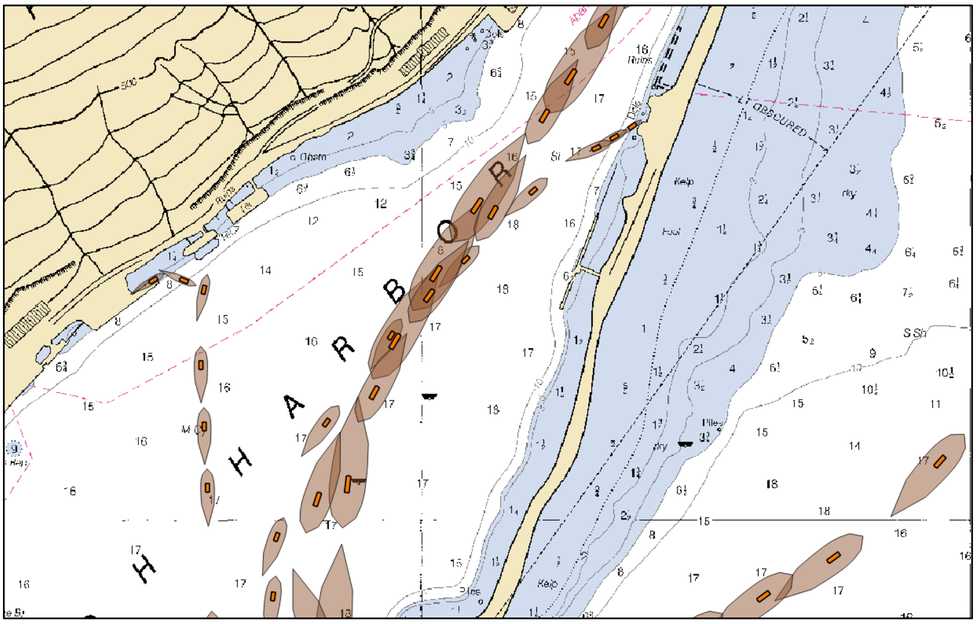


Figure 3: Example implementation of domain model (domains approximated with 18 vertices rather than true ovals).

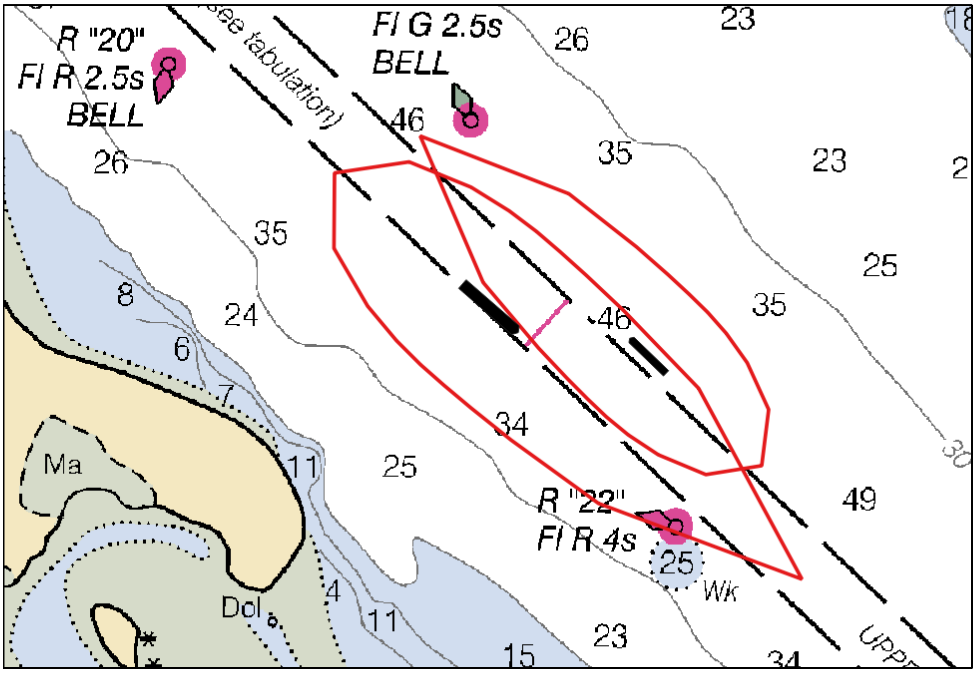


Figure 4:Example encounter in Columbia Channel of an inbound 218m vessel and an outbound of a 142m vessel.

3.3 ~~GENERATING~~ MEASURING ENCOUNTERS

In order to determine whether an encounter has occurred, the following workflow was developed, which is shown in Figure 5. Within the first part, a pairwise join of the processed AIS data is conducted such that at every minute, all vessels positions are compared. If the haversine distance between vessels is less than the calculated domain distance (described in Section 3.2) and the average speed of vessels is greater than 1 knot, an encounter is retained. This process removes stationary vessels such as those alongside that are not navigationally significant. Unlike previous studies which are limited to singular waterways and therefore able to use Euclidian Distances with metres based projection systems, our analysis covers a significant study area both latitudinally and longitudinally and therefore we are required to utilise haversine distances with the latitudes and longitudes of the vessel positions. The resulting encounter table contains numerous key characteristics about that encounter situation that are further analysed. This stage required an average of 2,800 vessels to be compared every minute for the full month of data and was therefore computationally expensive to run.

A second stage was necessary to process chains of consecutive encounters (i.e. occurring successively every minute during an overtaking situations), which we call Prolonged Encounters. We consider that a single record should be retained for each meeting situation between two vessels, and therefore a method to filter prolonged encounters is required. The encounter dataset is sorted by vessel MMSI numbers and encounter time, and if the encounter has the same vessel ids and occurred less than 10 minutes previously, the Prolonged Encounter ID number (PID) is kept the same. A 10-minute filter was utilised as occasional breaks between successive encounters occurred, before re-establishing the same chain. Finally, duplicate PID values were removed, retaining the encounter in each prolonged encounter that had the minimum separation between vessels.

For example, where two vessels encounter one another at a single timestamp, a single PID value is assigned which is unique in the dataset. If those two vessels were to encounter successively for five minutes, five encounter records are recorded, but each of those five records shares the same PID number. By removing duplicates, only the closest position of those five is retained, and similarly the single record in the first example is retained. The resulting dataset contained 348,543 encounters across the United States during June 2018.

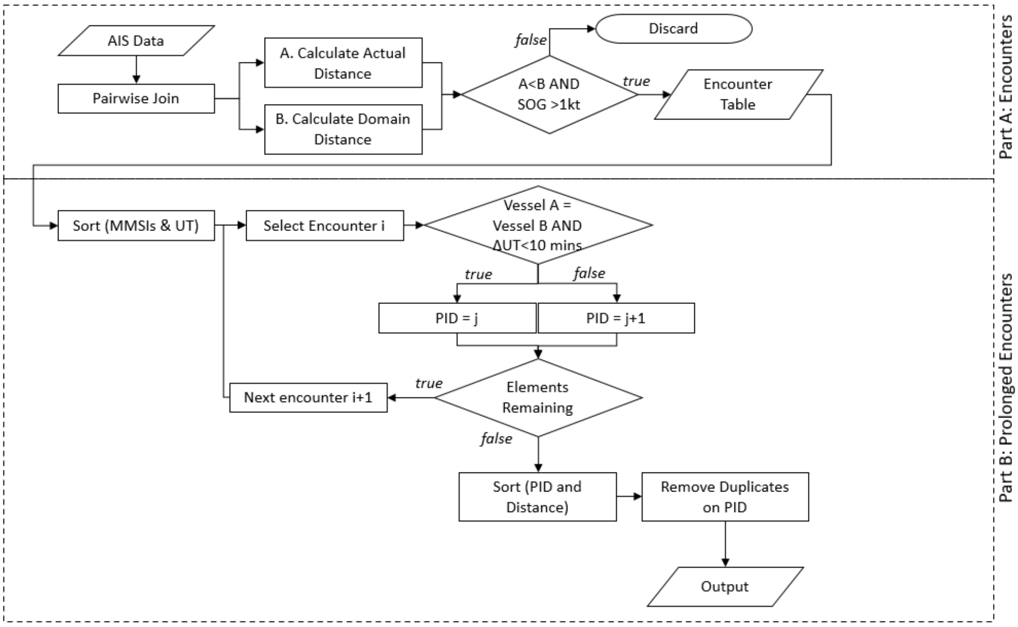


Figure 5: Encounter Processing Flowchart.

3.4 SPATIAL MODEL OF ENCOUNTERS AND INCIDENTS

In order to analyse the results and calculate correlations, a spatial data structure is required into which the data is ~~binned~~aggregated. One method would be through cartesian grids, however many authors have described inherent limitations of these formats when performing spatial analysis, such as distortion of shape, non-uniform adjacency between cells and variable cell areas when implemented at global or regional scales (Sahr and White, 1998). To overcome this, Discrete Global Grid Systems (DGGS) have been proposed, which partition the world into equal area platonic solids, such as hexagons or triangles (Sahr et al., 2003). In this case, we utilise hexagonal grids which have been shown to exhibit certain inherent advantages, such as uniform adjacency between cells and visual interpretability (Birch et al., 2007).

As part of the SEDNA project, the University of Southampton developed a python library that implemented the DGGRID R library (Barnes, 2016), called dggridpy (Correndo, 2019). The package enabled DGGS cells to be constructed at varying resolutions, and spatial data indexed. A hexagonal ISEA aperture 4 DGGS was developed at resolution 7, 8, 9 and 10 to support the analysis at different scales. Table 1 and Figure 6 compares the four resolutions assessed in this paper.

Table 1: Details of DGGS Resolutions

Resolution	Number of Cells (Global)	Cell Area (km^2)	Centroid Spacing (km)
7	21,872	23,322	151.2
8	65,612	7,774	82.3
9	196,832	2,591	50.4
10	590,492	863	27.3

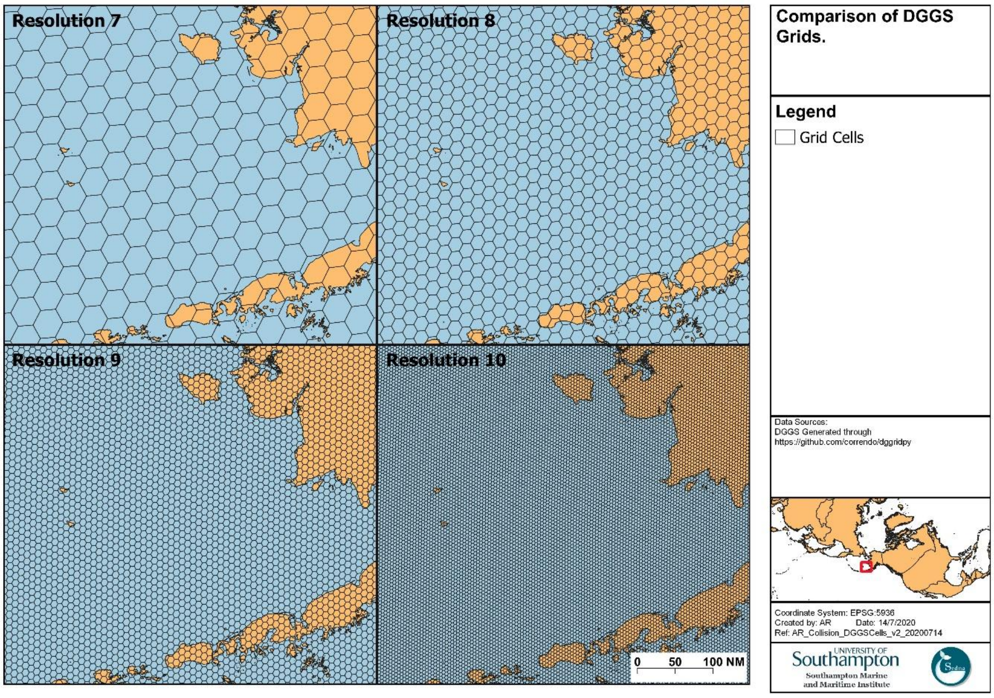


Figure 6: Comparison of DGGS Grids across Alaska.

3.5 FUTURE SCENARIO MODELLING

Whilst rarely considered in the literature, a key potential application of domain analysis is predicting the impact on collision risk of some future change in traffic. This might include an increase in the number of vessels, or alternatively, as taken in our example here, some future development, such as an offshore wind farm or new routeing measure. Such a development would require vessel traffic to divert around an obstacle, potentially creating choke points and increasing collision risk. Were the wind farm located elsewhere, with different geometries or traffic flows, the change in collision risk could be compared, potentially enabling optimisation of layouts to minimise the impact on collision risk. However, within this study we consider only a single development.

In such a circumstance, we can predict the expected number of collisions by multiplying the encounter frequency by a conditional probability (P_c) that an encounter would result in a collision ($P_{Collision} | P_{Encounter}$). We could measure the difference in collision risk by calculating the number of encounters under both the base and future case scenarios, provided the latter case can be sufficiently modelled.

$$\text{Collisions/Year} = Encounters/Year * P_c \qquad (6)$$

Figure 7~~Figure 7~~ shows a visualisation of a typical scenario which might be encountered. In this case, some physical obstruction is placed in the flow of traffic, which requires an alteration of a vessel's course to avoid the obstacle. Understanding how this scenario might change collision risk is not a trivial problem, and essential to forecasting the increase of risk of a proposed development or evaluating the requirement for risk control measures such as routeing schemes.

Formatted: Font: Bold

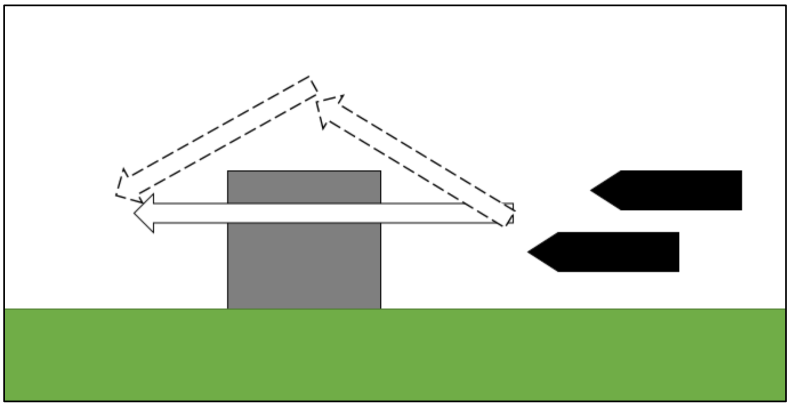


Figure 7: Dummy example of future scenario modelling.

~~3.5~~

Formatted: Normal

3.5.1 Case Study

~~Whilst rarely considered in the literature, a key potential application of domain analysis is predicting the impact on collision risk of some future change in traffic. This might include an increase in the number of vessels, or alternatively, as taken in our example here, some future development, such as an offshore wind farm or new routeing measure. Such a development would require vessel traffic to divert around an obstacle, potentially creating choke points and increasing collision risk. Were the wind farm located elsewhere, with different geometries or traffic flows, the change in collision risk could be compared, potentially enabling optimisation of layouts to minimise the impact on collision risk. However, within this study we consider only a single development.~~

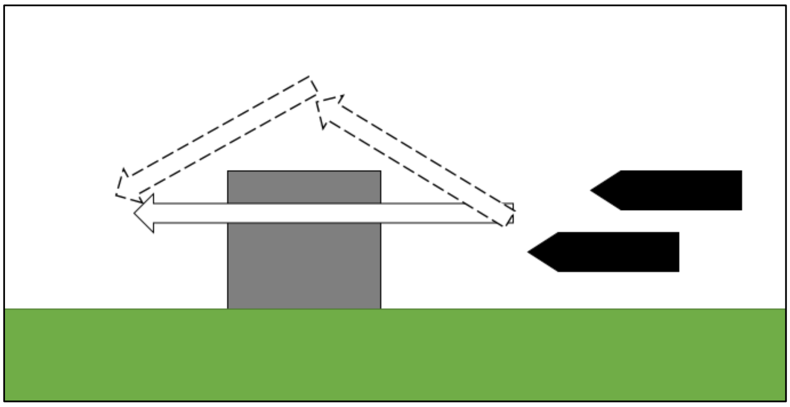
~~In such a circumstance, we can predict the expected number of collisions by multiplying the encounter frequency by a conditional probability (P_c) that an encounter would result in a collision ($P_{Collision}|P_{Encounter}$). We could measure the difference in collision risk by calculating the number of encounters under both the base and future case scenarios, provided the latter case can be sufficiently modelled.~~

Formatted: Subscript

$$\text{Collisions/Year} = Encounters/Year * P_c \qquad \cancel{(6)}$$

~~**Figure 7** shows a visualisation of a typical scenario which might be encountered. In this case, some physical obstruction is placed in the flow of traffic, which requires an alteration of a vessel's course to avoid the obstacle. Understanding how this scenario might change collision risk is not a trivial problem, and essential to forecasting the increase of risk of a proposed development or evaluating the requirement for risk control measures such as routeing schemes.~~

Field Code Changed



~~*Figure 7: Dummy example of future scenario modelling.*~~

To demonstrate this, we propose a hypothetical situation of an offshore wind farm to be constructed in the Straits of Juan de Fuca, Washington State (**Figure 8**~~**Figure 8**~~). The Straits are the major approach routes between the continental USA and Vancouver Island, serving several major ports such as Vancouver, Seattle, Everett and Victoria. The route has a ~~Traffic Separation Scheme (~~TSS~~)~~, which mandates single direction travel for large commercial vessels. In our situation, the hypothetical wind farm boundary extends to the limits of the southern (inbound) traffic lane.

Formatted: Font: Bold

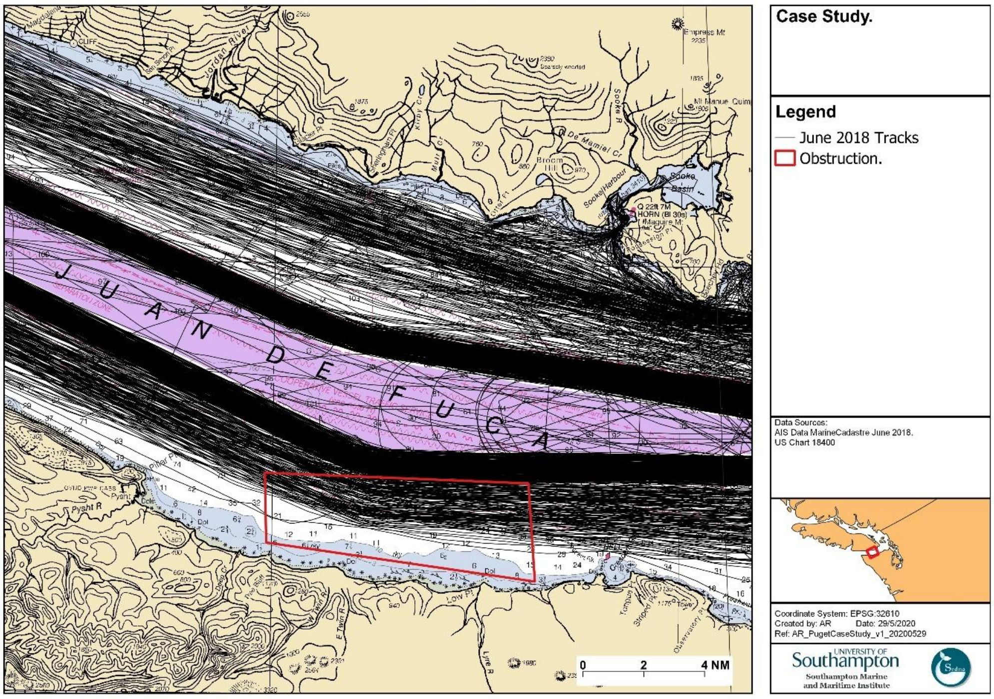


Figure 8: Vessel traffic in the Straits of Juan de Fuca and hypothetical wind farm (June 2018).

In order to test how this development might change collision risk, we must first calculate the baseline number of encounters, as described in the preceding sections. Following this, a method of re-routeing traffic flow is required, and then the domain model re-tested with the new traffic flows to measure the increase.

3.5.2 Re-Routeing Vessel Traffic

Numerous methods are available for routeing vessel traffic, including the use of path finding algorithms such as A* or Dijkstras (Ari et al., 2013). In this case study, a simple divers~~is~~ion method is utilised, which is described in Figure 9 below. The method takes tracks that pass through an obstruction and assigns new coordinates related to boundary markers that indicate the expected course for vessels around the obstruction. The process is to:

1. Identify vessel trips, taken by the same vessel, such that the subsequent positions are less than 10 minutes apart. If not, create new trips.

2. Spatial query as to whether a trip passes through the obstruction area.
    a. If not (i.e. transit A in Figure 9), set aside.
    b. If so:
        i. Determine the closest boundary coordinates to each position using a KDTree nearest neighbour search.
        ii. Drop duplicates for each trip ID and boundary coordinate.
        iii. For each trip, assign a random track offset using a normal distribution with standard deviation 100m.
        iv. Create new coordinates using boundary coordinates and track offsets.
3. Merge diverted and un-diverted positions to new database.

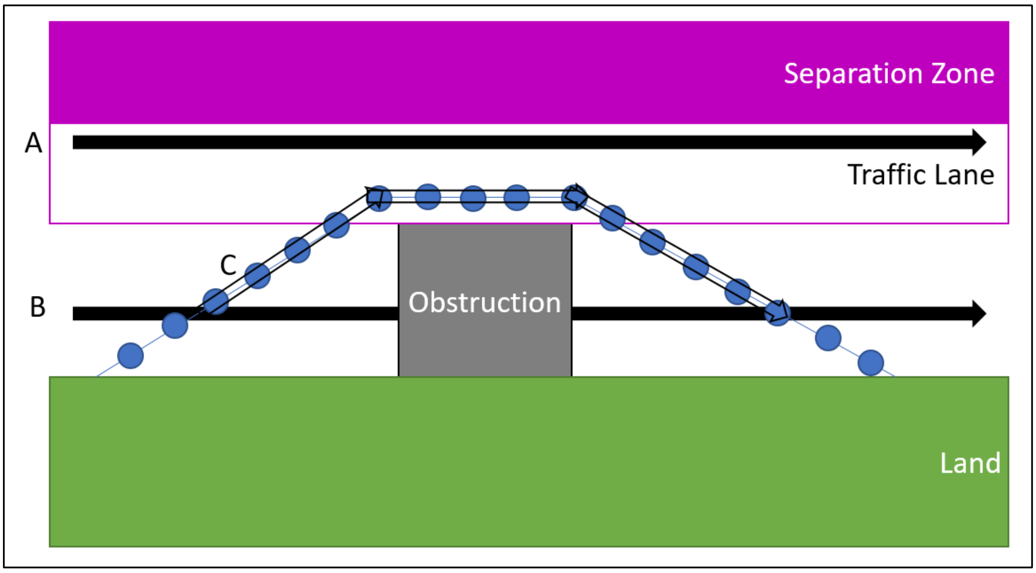


Figure 9: Route diversion method.

Figure 10~~Figure 10~~ shows the results of this on the June 2018 dataset. Several notes should be made on the applicability of this method in real scenarios as opposed to our hypothetical one. Firstly, the method has maintained transit in the traffic lane only, hence there is a significant concentration of traffic to the north of the obstruction which would be unrealistic. In practice, it would be likely that the traffic lanes would be reconfigured to maintain sufficient separation from the obstruction. Secondly, we have assumed that vessels transiting outside of the lanes normally would continue to transit in this fashion except when passing the obstruction. This results in significant alterations of course for some vessels that are close to the southern shoreline. Thirdly, we have not included any

Formatted: Font: Bold

alteration of transit time, assuming that vessel speed is increased to cover the increased distance. Given the relatively small increased transit distance in this example, this is not considered a significant limitation, but may be much more pronounced in other applications~~. Therefore, further work is required to improve the ship diversion method~~ and therefore warrants further work.

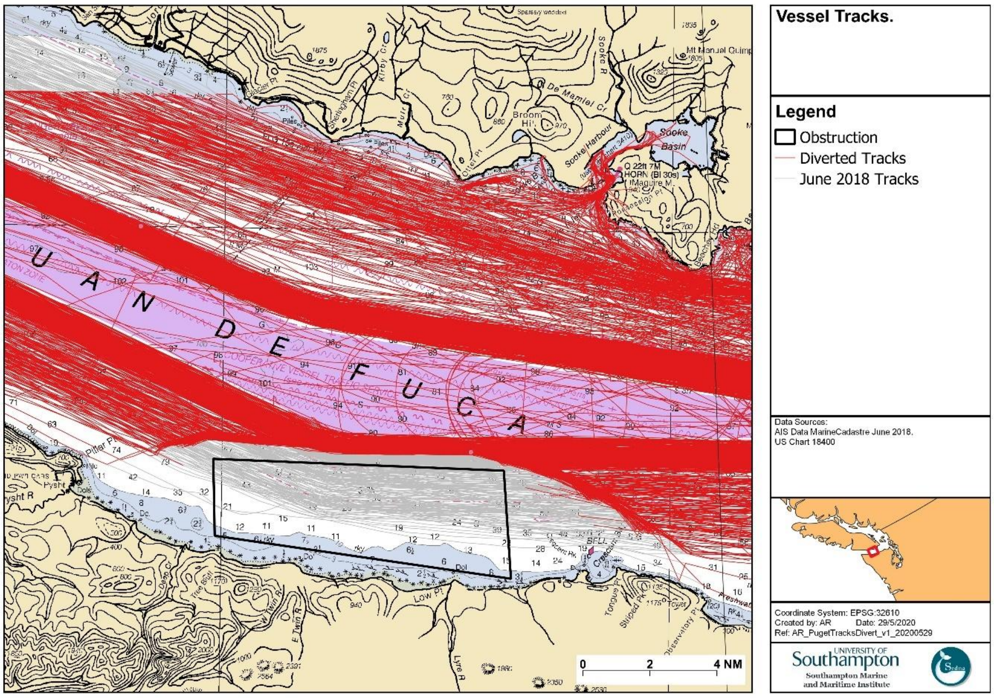


Figure 10: Diverted vessel traffic.

4 RESULTS

4

Formatted: Normal

4.1 IMPLEMENTATION OF DOMAIN ANALYSIS

The June 2018 AIS data was processed and the number of encounters, scaled to annual figures, in each location mapped in **Figure 11**~~Figure 11~~ and **Figure 12**~~Figure 12~~. For much of the county, the number of encounters is low, but where vessel traffic is concentrated, such as the approaches to ports, the frequency is increased. **Figure 12**~~Figure 12~~ compares the number of encounters and historical number of collisions in the ~~the~~ Gulf of Mexico, with the results suggesting a relationship

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Font: Bold

does exist. It is evident that mapping the distribution in encounters can show differences in traffic disposition between locations, but we must then test the strength of the relationship between encounters, and collision frequency.

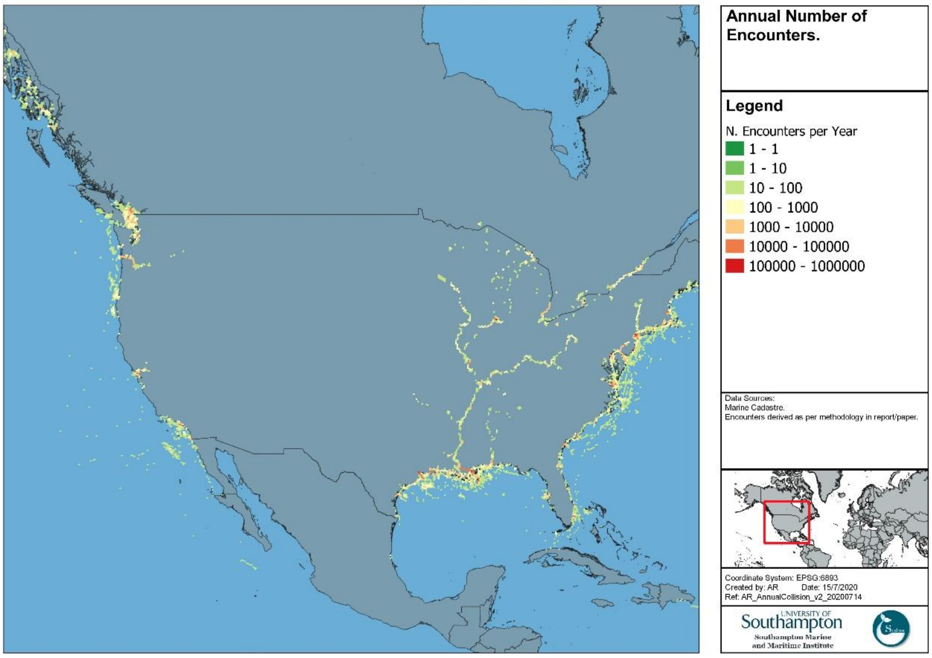


Figure 11: DGGS9 Encounter Rates.

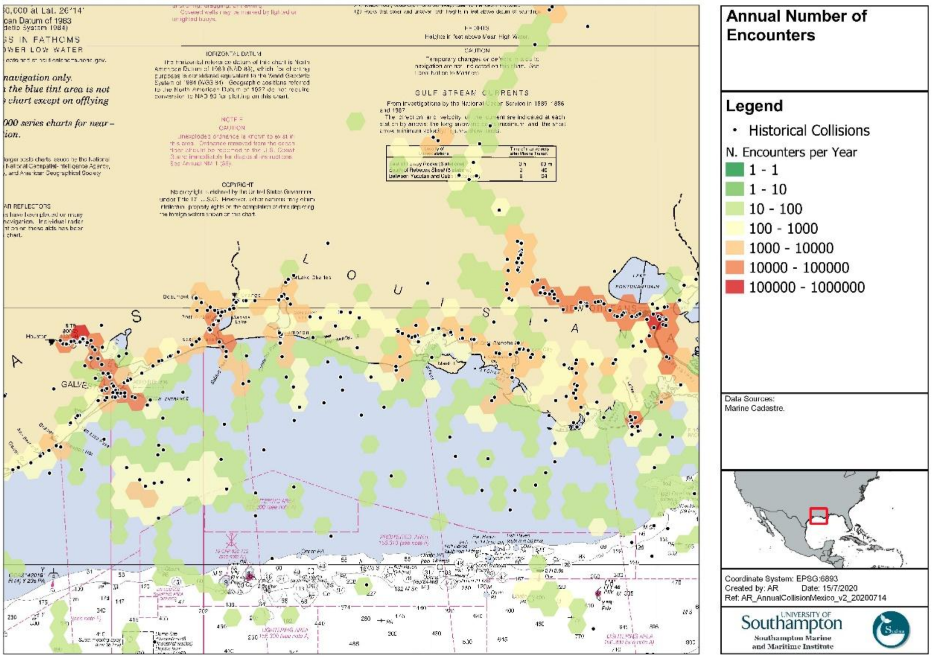


Figure 12: DGGS9 Encounter Rates and Historical Collisions in Gulf of Mexico.

4.2 RELATIONSHIP BETWEEN ENCOUNTERS AND COLLISIONS

The Pearson R (SciPy, 2020) values across a number of spatial grid resolutions and for each vessel type are shown in **Table 2**~~Table 2~~. Firstly, in agreement with previous research the results are subject to the Modifiable Areal Unit Problem (MAUP), that results in the correlations increasing as the cell sizes become larger and the results more generalised (Rawson et al. 2019). The resolution at DGGS9 was taken forward as a compromise between high spatial resolution but maintaining good correlation and the relationship at this resolution is shown in **Figure 13**~~Figure 13~~.

Formatted: Font: Bold

Formatted: Font: Bold

Formatted: Caption,~Caption, Justified, Space After: 0 pt, Line spacing: single, Keep with next

Table 2: Pearson R Values (indicates p value greater than 0.05).*

	Encounters				Exposure	Variants	
	DGGS7	DGGS8	DGGS9	DGGS10	DGGS9	DGGS9 (>1nm)	DGGS9 (>10nm)
Total	0.63	0.57	0.53	0.36	0.66	0.50	0.06
Commercial	0.71	0.65	0.61	0.52	0.33	0.63	0.07
Fishing	0.27	0.22	0.18	0.15	0.18	0.11	0.01
Passenger	0.32	0.32	0.28	0.27	0.38	0.30	0.01
Recreational	0.54	0.51	0.24	0.27	0.41	0.38	0.01*
Other	0.64	0.58	0.54	0.33	0.71	0.24	0.01*

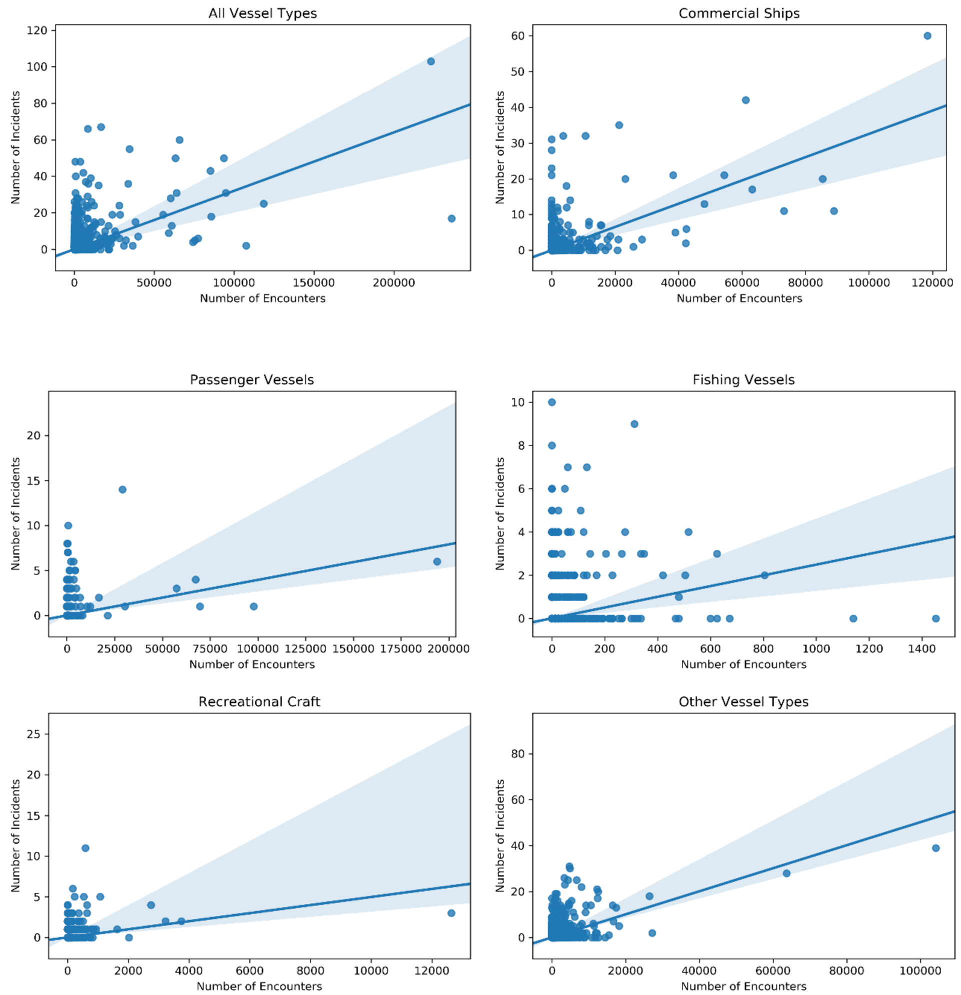


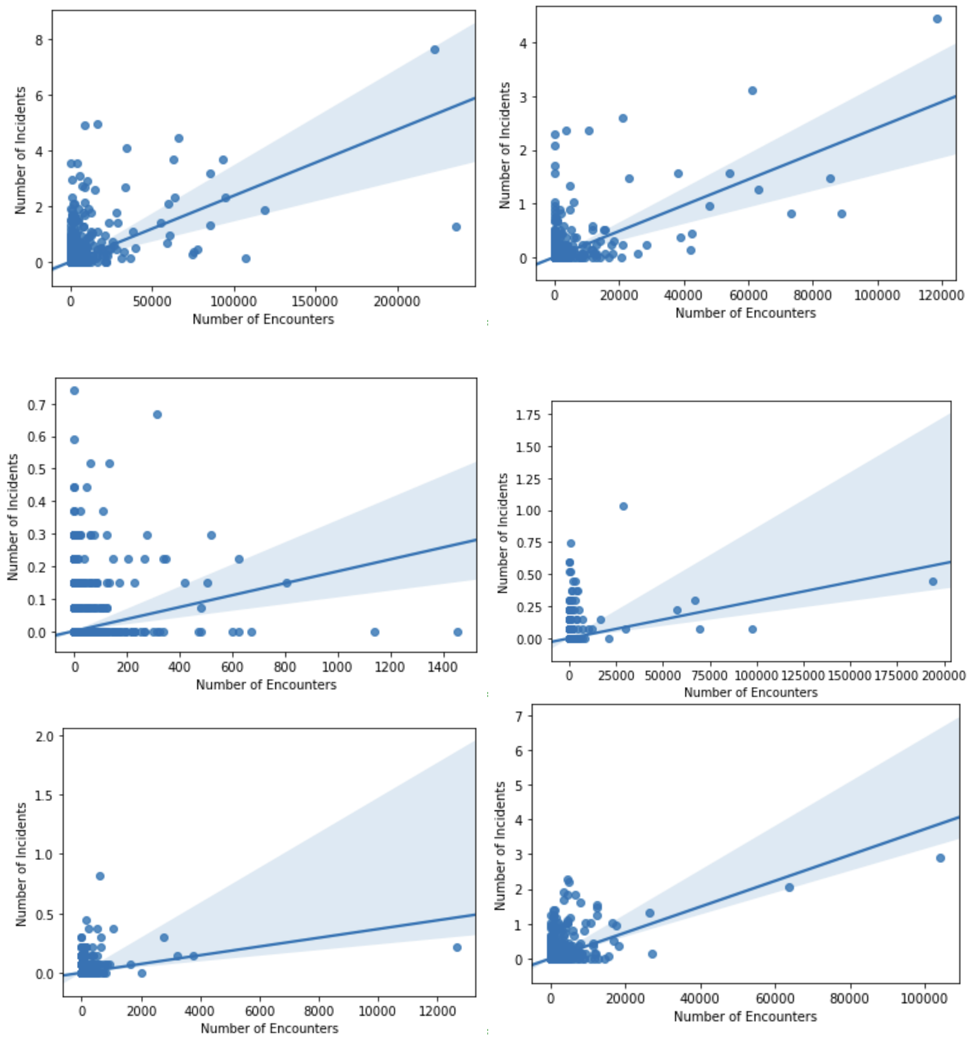
Figure 13: Relationship between encounters and collisions at DGGS Resolution 9.

Secondly, the results are varied between vessel types, with commercial having the greatest at DGGS9 of 0.61 and fishing vessels having the weakest at 0.18. Other vessel types, such as tug and tows and dredgers have similar correlations, exceeding that of passenger and recreational vessel types. The relative low correlation of fishing and recreational maybe partly explained by the low carriage of AIS, which might be disproportionately greater in inshore areas where encounters are more likely to occur than offshore, where traffic is generally less concentrated. In addition, some fishing activities require multiple vessels to operate in close proximity, generating encounters but not necessarily increasing collision risk. There is also a question of incident data completeness, that might result in higher reporting rates for commercial or passenger vessels than fishing or recreational vessels.

Thirdly, the results are compared against a simple count of exposure (hours of operation per grid cell) at the same resolution. Interestingly, for some vessel types this had a higher correlation for historical collisions than the number of encounters. The results suggest that passenger, recreational and other vessel collisions can be better predicted by vessel exposure, although commercial vessel collisions are better described using encounters. This suggests that simple methods of risk assessment, such as using simple measures of traffic exposure, can achieve near equivalent results, without the need for significant computation.

Finally, sensitivity testing was conducted at DGGS9 resolution by including only those grid cells that were more than 1nm and 10nm from the coast, respectively. At 1nm, which includes coastal cells, but excludes inshore cells (such as the Mississippi River), the correlations are largely unchained. However, at 10nm, the correlations reduce significantly. This might be partly due to reduced coverage of AIS at greater distance from shore, reduced sample size of collisions, or indicate a possible difference in navigation behaviour offshore as opposed to coastal. Whilst this latter point is logical, vessels navigating offshore have more space and time to plan avoiding manoeuvres than in constrained waterways, due to the aforementioned limitations we are unable to conclusively demonstrate a difference in navigation behaviour between these settings.

~~***Table 2: Pearson R Values (* indicates p-value greater than 0.05).***~~



~~Figure 13: Relationship between encounters and collisions at DGGS Resolution 9. From Top left: Total, Commercial, Fishing, Passenger, Recreational and Other.~~

Furthermore, using the analysed data, it is possible to compare the relationship between encounters and exposure (Figure 14). In areas with significant numbers of movements, it would be expected that vessels would inevitably navigate closer together. For most vessel types the results support this, yet fishing vessels show a significant number of low exposure and high encounter locations, likely situations where fishing vessels are working equipment in groups. This further supports the finding that vessel traffic exposure may be equivalent to vessel encounters as a predictor of collision frequency. In Rawson et al. (2014) it was shown that the River Thames exhibited a non-linear relationship, as the river became busier with more activity, the number of encounters increased exponentially. It is likely that different environments will have different relationships between these variables, for example as a function of the waterway geometry, and therefore further research is necessary using different variables or statistical models to better describe this relationship.

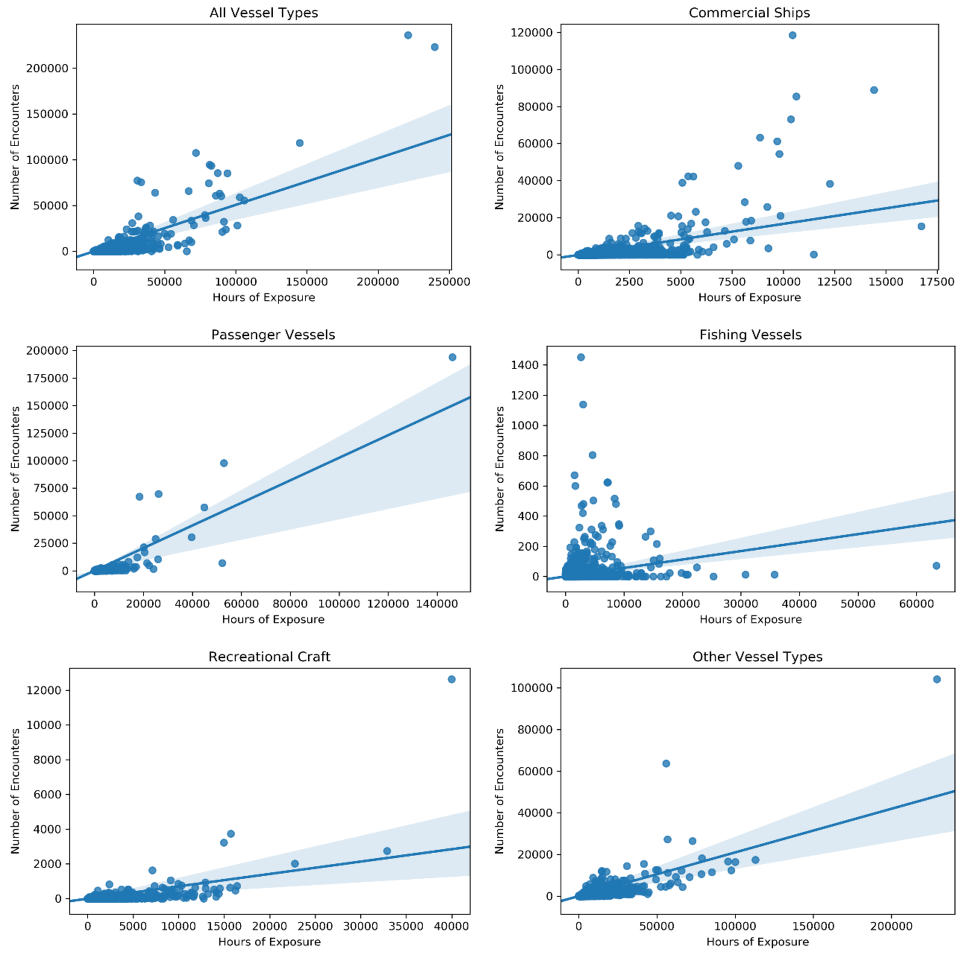


Figure 14: Relationship between encounters and exposure at DGGS Resolution 9.

Formatted: Caption,~Caption, Left

4.3 FUTURE SCENARIO MODELLING

Finally, the results of the future scenario model are described in this section. Based on the re-routeing of traffic conducted, the number of encounters before and after the development can be established. Table 3 shows these results, with the re-routeing resulting in a 234% increase in the absolute number of encounters, and a 778% increase in unique encounter situations. If we annualise the number of encounters and apply a causation probability of ($P_{Collision}|P_{Encounter}$) where Pc is equal to 4.9×10^{-5} (Goerlandt and Kujala, 2014), then annual collision probabilities can be obtained. The results show an increase from once in 189 years to once in 21.5 years.

These encounters are shown in Figure 15, and are clustered adjacent to the obstruction, where traffic flow is most concentrated. This methodology enables quantification of the magnitude of the impact of the development on collision risk, and presentation of the spatial distribution of that risk, in order to target risk controls effectively. A discussion of some of the key assumptions in this approach is contained in Section 5.3.

Table 3: Results

Model	Encounters (June18)	Unique Encounters (June18)	Unique Annualised	Causation Probability (Pc)	Collisions/ Year
Baseline	166	9	108	4.9×10^{-5}	5.29×10^{-3}
Re-Routed	553	79	948	4.9×10^{-5}	4.65×10^{-2}

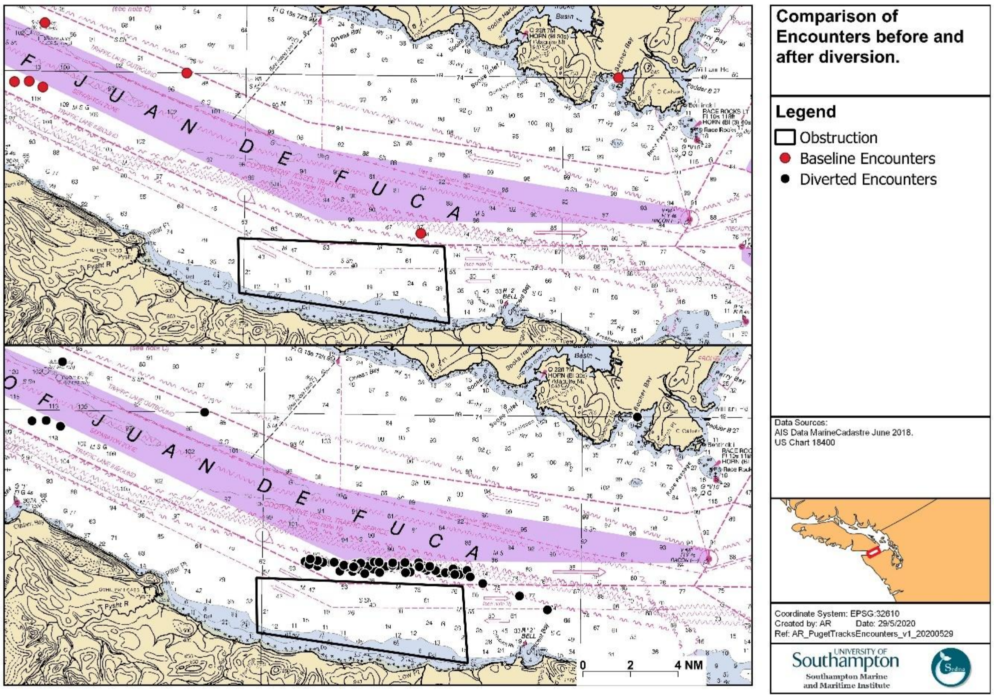


Figure 15: Comparison of encounters between models.

5 DISCUSSION AND FUTURE WORK

~~5~~ Within our discussion we consider three key findings of our work. Firstly, we consider the statistical relationship between the number of encounters and collisions and the implications of this result on collision risk modelling. Secondly, we consider how specific implementations of ship domains might impact the results, and whether some models are more suited to one environment than another. Finally, we consider some of the key methodological aspects of our future scenario modelling, highlighting some limitations that warrant further investigation.

Formatted: Normal

5.1 ENCOUNTERS AS A PREDICTOR FOR COLLISIONS

The results are mixed in the predictive power of encounters and collision risk, with Pearson correlations between 0.61 and 0.18. This result is perhaps not surprising given the inherent complexity in predicting accidents in a highly complex system, with numerous inter-relating contributory factors (Du et al. 2020). In addition, the high contribution of human factors to historical collisions is well known, which cannot be directly observed from AIS data. Other external factors which might influence this relationship are weather conditions, waterway geometry, hydrodynamic effects and the presence of risk controls such as pilotage (Mazaheri et al. 2014; Bye and Aalberg, 2018; Olba et al. 2019). Further work is proposed to better understand whether the inclusion of these variables when modelling collision risk achieves better results.

In addition, the implementation of this approach treats encounters as pairwise interactions between vessels, which may result in a collision. As a result, it fails to capture complex multi-vessel encounter situations, which might have a greater likelihood of resulting in a collision and can only be evaluated using alternative methods (Zhang et al. 2020).

The quality of data should also be considered. The MarineCadastre does not have complete AIS coverage of all areas of the US coastline, therefore where this coverage is worst, there would be a relatively low number of encounters given reported collisions. Similarly, the further offshore, AIS reception degrades and a similar result would be expected. This might also apply for certain vessel types, with fishing and recreational having lower AIS carriage rates than commercial shipping. In addition, the relative under-reporting of incidents is well known (Qu et al 2012; Hassel et al. 2011), which makes establishing correlations difficult, particularly for non-commercial vessels.

~~Some authors have presented non-binary domain models where the encounter is weighted by the significance of the encounter, with high speed and close encounters assigned more weight than slow speed and distant encounters. Whilst we have not included this aspect in our chosen implementation of Wang's (2010) model, we do not believe this would significantly alter the spatial distribution of the results. An interesting point is raised here as to whether such an approach adds any value to strategic use of domain analysis in collision risk assessment, given that the size and shapes of the domains already vary based on the size and speed of the vessel. The authors are not aware of any work that concludes that non-binary domain encounters are better predictors to the frequency of collisions than binary domain encounters.~~

Whilst the correlations of the models are varied, its implementation in characterising where collisions are more likely shows that collision risk can be characterised across regions. This supports decision makers in ~~correctly~~ determining where risk is highest and implementing targeted risk controls to reduce that risk. Yet, our results also show that vessel traffic exposure can achieve almost equivalent results in some vessel types. Some relationship between vessel traffic activity and the number of activities is to be expected, yet the strength of this relationship suggests that simple measures of vessel movements can be used to characterise collision risk. This approach has numerous advantages as it requires less technical input, less computation and therefore has greater utility for navigation authorities.

Theoretically, we consider that the principal advantage of domain modelling over other models (such as geometric route intersections) is the inclusion of the time element. Therefore, a domain model would better represent a system where there is significant variability in spatiotemporal activity, such as where a day/night or tidal aspect is significant. In addition, in our future case scenario where the underlying vessel traffic volume remains unchanged, domain analysis is able to show not only a quantitative increase in the number of encounters, but where those additional encounters occur. This provides much more targeted information than traffic volume or incident rates could provide.

5.2 SUITABILITY OF DOMAIN MODEL

As a result of the multitude of domain shapes available for implementation by an analyst seeking to predict collision risk (Fiskin et al. 2020), it is likely that domain shape has some influence on the results (Goerlandt and Kujala, 2014). Given that there is no agreed ship domain concept, some limitations are inevitable on our adoption of Wang (2010), and these are briefly commented upon below.

Firstly, as has have been summarised by Zhang et al. (2016) there are certain limitations with oval domain shapes more generally. Ovals are symmetrical and therefore suggests that the influence of the COLREGs, which take into account the direction that meeting vessels should cross, is asymmetrical, and previous empirical analysis of ship domains has shown this (Hansen, 2013). Where the domain extends behind a vessel, particularly in Fujii and Tanaka (1971) where the forward and aft domains are equal, this suggests that crossing astern of a vessel is equal to crossing ahead of a vessel. In some cases, including in our implementation, it is possible for one vessel to encounter another, but not vice versa. For example, a large ship with a large domain would encounter a passing fishing vessel, but due to their much smaller domain, they would not encounter the ship. In this situation, we record an encounter, but some implementations may suggest that the risk of collision lies with the ship only and not the fishing vessel, when clearly one would necessarily collide with the other.

Some authors have presented non-binary domain models where the encounter is weighted by the significance of the encounter, with high speed and close encounters assigned more weight than slow speed and distant encounters. Whilst we have not included this aspect in our chosen implementation of Wang's (2010) model, we do not believe this would significantly alter the spatial distribution of the results. We expect instead that congested waterways where ships come close together would have both more encounters and those encounters would naturally be closer together on average. Conversely open waters enable masters to increase their passing distance, resulting in less encounters and at a greater distance on average. An interesting point is raised here as to whether such an approach adds any value to strategic use of domain analysis in collision risk assessment, given that the size and shapes of the domains already vary based on the size and speed of the vessel. The authors are not aware of any work that concludes that non-binary domain encounters are better predictors to the frequency of collisions than binary domain encounters, but this could be investigated as part of future work.

The formulas provided by many authors fail to account for many environmental and circumstantial details which might alter the natural separation between vessels. Reflective of this are the great number of studies that have focused on applications within a single waterway, in order to demonstrate their proposed domain shape. For example, Rawson et al. (2014) propose a shape fitted to navigation on the River Thames, which is necessarily different to ones proposed for open sea (Fujii and Tanaka, 1971), port approaches (Bakdi et al. 2019; Fang et al. 2019) or TSS (Zhang and Meng, 2019). The influence of waterway characteristics is therefore a major determining factor of navigation behaviour and therefore domain shape (Szlapczynski and Szlapczynska~~i~~, 2017). Some have argued that a universal ship domain might be achievable (Pietrzykowski and Uriasz, 2009), applicable to all waterways, but this may not be possible or even desirable if they are applied for specific purposes.

This effect may be important, in constrained waterways where vessels naturally navigate closer together. Should the size of the domain reduce to account for the routine proximity at which vessels operate to differentiate normal encounters to abnormally close encounters. Similarly, it could be argued that as vessels navigate close together, bridge teams compensate by increased alertness, maybe even navigating with pilots, and therefore the awareness and reaction capability is far quicker, supporting smaller domains than at open sea. By contrast, more collisions occur in ports and constrained waters, supporting the association between collisions and encounters. Further work is planned in order to investigate the impact of waterway characteristics on empirically derived domain shapes.

5.3 DETERMINING INCREASED COLLISION RISK IN FUTURE SCENARIOS

As described in Section 4.3, by re-routeing vessel traffic around a proposed development, we can predict how the number of encounters might increase. Whilst we have raised questions on the relationship between number of encounters and historical collisions; in this situation, given that the traffic volume between the base and future case situations is the same, domain analysis has the advantage of being able to characterise how risk changes, something that traffic volume alone could not achieve.

An outstanding issue is the degree to which the~~is~~ increase in encounters is reflective of an increase in collision risk, rather than artificial encounters as a result of the modelling methodology. For example, if we take the number of encounters as E, then the increase in encounters takes the form:

$$\Delta E = E_{Modelled} - E_{Baseline}$$ (~~6~~7)

In the baseline case, vessels are avoiding one another and attempting to maintain domain separation. Once traffic is re-routed, the ships may be assigned new paths which bring them into conflicts with other navigating vessels. Whilst this is useful, as it indicates that two vessels would now meet when they would not previously, there is an unknown likelihood that one of those vessels would have taken action to increase the separation distance, thereby avoiding an encounter. Some factor P_A, the probability of avoidance, is necessary to adjust the resulting increases in encounter situations.

It should be noted that this is distinct from the P_C or causation probability used in other risk assessment methods, which is the conditional probability of a collision given an encounter ($P_{Collision}|P_{Encounter}$), which might be in the region of 4.9×10^{-5} (Goerlandt and Kujala, 2014). Therefore, we should utilise the equation in the form to correct for this:

$$\text{Collisions} = (((E_{Modelled} - E_{Baseline}) * P_A) + E_{Baseline}) * P_c$$ (~~7~~8)

This aspect is also relevant~~,~~ were some risk algorithm incorporated into the equation, such that collision risk is calculated per encounter. As proposed by some authors, inclusion of vessel speed, size and type can graduate the severity of the encounter situations. Yet, re-routed vessels following blind navigation may overlap with existing traffic, and have high speed, zero-distance encounters, that would generate artificially high collision risk scores.

P_A is an unknown factor, but some proposed methods for solving this are discussed below. Firstly, this could be estimated using expert judgement based on the perceived probability of avoiding another approaching vessel. Secondly, by comparing the rate of encounters to transits in other waterways which have similar characteristics as that of the diverted scenario, may provide some indication of the chance of two vessels encountering in compressed waterways. Finally, the baseline scenario could itself be modelled, by generating the baseline traffic along a route, the difference between the actual baseline and the modelled baseline reflects new encounters ~~as a result~~ ~~o~~generated due to ~~f~~ the modelling alone.

A secondary point when applying this methodology is the size of the study area. In our case study, we utilised only a small section of the Strait of Juan de Fuca, and therefore the impact on collision risk was localised. As a result, the relative increase in risk in this small section is significant. If the results are replicated for the whole Strait, then the relative increase would be much less. This would impact upon stakeholder reception of the analysis results, presenting high or low relative increases might influence their decision making of what is acceptable. This therefore requires some degree of presentation of risk in both absolute and relative terms, which domain analysis would facilitate.

6 CONCLUSIONS

This paper has investigated the practical applications of domain analysis in strategic risk assessments and critiqued their utility at predicting present day and future case collision risk. Whilst other methods of collision risk assessment exist, the widespread adoption of domain analysis within both academia and industry warrants a detailed evaluation to ensure it is fit for purpose. In testing this, our work has achieved the following key findings. Firstly, the statistical relationship between encounters and collisions is not strong, particularly for certain vessel types. This suggests that, whilst this approach may be useful in determining where collision risk is highest, ~~The results suggest that domain analysis remains a useful tool to predict the number of collisions, but that the correlation between encounters and collisions may not be strong for certain vessel types and certain environments. As a result,~~ the inclusion of other variables such as environmental conditions and local practice may be necessary to characterise the risk more accurately. This has been attempted by some authors, but the improvement in the predictive accuracy of these approaches has not been demonstrated. Secondly, as with other spatial models where data is aggregated, the statistical relationship is impacted by the MAUP, whereby the scale of assessment impacts the statistical relationships derived. Such an effect must therefore be considered by authors when employing this approach. Thirdly, we demonstrate that using a measure of vessel activity has predictive accuracy equivalent to frequency of encounters in some cases as a predictor of collision frequency. This suggests that simple methods, such as volume of traffic, may be as accurate as more complex methods for use in maritime risk assessment, with significantly less computational requirements.

These findings aside, in addition

~~Furthermore,~~ we have demonstrated the applicability of domain analysis to quantify the impact of a development or obstacle on collision risk. This supports evidence-based decision making on the magnitude of risks, and the effectiveness of proposed mitigation, during the permitting application and environmental studies of projects such as wind farms. Yet, several key challenges with this approach are highlighted that require additional research. Most notably the action of diverting the vessel traffic to avoid an obstruction~~,~~ generates a significant number of encounters which unrealistically do not avoid one another. Resolving this limitation is essential before domain analysis's strong capability to predict changes in risk can be realised.

ACKNOWLEDGEMENTS

This work is partly funded by the University of Southampton's Marine and Maritime Institute (SMMI) and the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement number: 723526: SEDNA).

REFERENCES

Anatec, 2019. Thanet Offshore Wind Farm: Collision Assessment of Proposed Extension. https://infrastructure.planninginspectorate.gov.uk/wp-content/ipc/uploads/projects/EN010084/EN010084-001987-Vattenfall%20Wind%20Power%20Limited%20-%20D6_Appendix42_TEOW_CRM_RevA.pdf (accessed 18 February 2020).

Ari, I., Aksakalli, V., Aydogdu, V., Kum, S., 2013. Optimal Ship Navigation with Safety Distance and Realistic Turn Constraints. European Journal of Operational Research 229(3), 707-717. https://doi.org/10.1016/j.ejor.2013.03.022.

Aven, T., Heide, B. 2009. Reliability and validity of risk analysis. Reliability Engineering and System Safety 94, 1862-1868. https://doi.org/10.1016/j.ress.2009.06.003.

Bakdi, A., Glad, I., Vanem, E., Engelhardtsen, O., 2019. AIS-based Multiple Vessel Collision and Grounding Risk Identification based on Adaptive Safety Domain. Journal of Marine Science and Engineering 8, 5. https://doi.org/10.3390/jmse8010005.

Barnes, R., 2016. dggridR: Discrete Global Grid Systems for R. https://github.com/r-barnes/dggridR (accessed 24 May 2019).

Bye, R., Aalberg, A., 2018. Maritime navigation accidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports. Reliability Engineering and System Safety, 176, 174-186. https://doi.org/10.1016/j.ress.2018.03.033.

Bye, R., Almklov, P., 2019. Normalization of maritime accident data using AIS. Marine Policy 109, 103675. https://doi.org/10.1016/j.marpol.2019.103675.

Chen, P., Huang, Y., Mou, J., van Gelder, P., 2019. Probabilistic risk analysis for ship-ship collision: State of the art. Safety Science 117, 108-122. https://doi.org/10.1016/j.ssci.2019.04.014.

Correndo, G., 2019. DGGRIDPY Git-hub page. https://github.com/correndo/dggridpy (accessed 22 May 2020).

Du, L., Goerlandt, F., Kujala, P., 2020. Review and analysis of methods for assessing maritime waterway risk based on non-accident critical events detected from AIS data. Reliability Engineering and System Safety 200, 106933. https://doi.org/10.1016/j.ress.2020.106933.

EMSA, 2019. Annual Overview of Marine Casualties and Incidents 2019. http://www.emsa.europa.eu/emsa-homepage/2-news-a-press-centre/news/3734-annual-overview-of-marine-casualties-and-incidents-2019.html. (accessed 18 May 2020).

Fang, Z., Yu, H., Ke, R., Shaw, S., Peng, G., 2019. Automatic Identification System-Based Approach for Assessing the Near-Miss Collision Risk Dynamics of Ships in Ports. IEEE Transactions on Intelligent Transportation Systems 20(2), 534-543. https://doi.org/10.1109/TITS.2018.2816122.

Feng, Z., Yang, H., Li, X., Li, Y., Liu, Z., Liu, R., 2019. Real-time Vessel Trajectory Data-Based Collision Risk Assessment in Crowded Inland Waterways. 4th IEEE International Conference on Big Data Analytics, 128-134. https://doi.org/10.1109/ICBDA.2019.8712843.

Fiskin, R., Nasiboglu, E., Yardimci, M., 2020. A knowledge-based framework for two-dimensional (2D) asymmetrical polygonal ship domain. Ocean Engineering, 202. https://doi.org/10.1016/j.oceaneng.2020.107187.

Fournier, M., Hilliard, C., Rezaee, S., Pelot, R., 2018. Past, present and future of the satellite-based automatic identification system: areas of applications (2004-2016). WMU Journal of Maritime Affairs 17, 311-345. https://doi.org/10.1007/s13437-018-0151-6.

Fujii, Y., Tanaka, R., 1971. Traffic Capacity. Journal of Navigation 24, 543-552. https://doi.org/10.1017/S0373463300022384.

Goerlandt, F., Kujala, P., 2014. On the reliability and validity of ship-ship collision risk analysis in light of different perspectives on risk. Safety Science 62, 348-365. https://doi.org/10.1016/j.ssci.2013.09.010.

Goodwin, E. 1975. A statistical study of ship domains. Journal of Navigation 28, 328-344. https://doi.org/10.1017/S0373463300041230.

Hanninen, M., 2014. Bayesian Networks for Maritime Traffic Accident Prevention: Benefits and Challenges. Accident Analysis and Prevention 73, 305-312. https://doi.org/10.1016/j.aap.2014.09.017.

Hansen, M.G., Jensen, T.K., Lehn-Schiøler, T., Melchild, K., Rasmussen, F.M., Ennemark, F., 2013. Empirical Ship Domain based on AIS Data. Journal of Navigation 6, 931-940. https://doi.org/10.1017/S0373463313000489.

Harati-Mokhtari, A., Brooks, P., Wall, A., and Wang, J. (2007). Automatic Identification System (AIS): Data Reliability and Human Error Implications. Journal of Navigation 60, 373-389. https://doi.org/10.1017/S0373463307004298.

Hassel, M., Asbjornslett, B., Hole, L., 2011. Underreporting of maritime accidents to vessel accident databases. Accident Analysis and Prevention 43, 2053-2063. https://doi.org/10.1016/j.aap.2011.05.027.

Heinrich, H. 1931. Industrial Accident Prevention. A Scientific Approach. McGraw-Hill Book Company, Inc. New York.

IALA, 2002. IALA Guidelines on the Universal Automatic Identification System (AIS). Volume 1, Part II – Technical Issues. Edition 1.1.

Im, N., Luong, T., 2019. Potential risk ship domain as a danger criterion for real-time ship collision risk evaluation. Ocean Engineering 194, 106610. https://doi.org/10.1016/j.oceaneng.2019.106610.

Kulkarni, K., Goerlandt, F., Li, J., Banda, O., Kujala, P., 2020. Preventing shipping accidents: Past, present, and future of waterway risk management with Baltic Sea focus. Safety Science 129, 104798. https://doi.org/10.1016/j.ssci.2020.104798.

Li, S., Meng, Q., Qu, X., 2012. An Overview of Maritime Waterway Quantitative Risk Assessment Models. Risk Analysis 32(3), 496-512. https://doi.org/10.1111/j.1539-6924.2011.01697.x.

MAIB, 2014. Report on the investigation of the collision between Paula C and Darya Gayatri in the South-west lane of the Dover Traffic Separation Scheme on 11 December 2013. Report No 25/2014.

Marico Marine, 2018. Thanet Extension Offshore Wind Farm: Navigation Risk Assessment. https://infrastructure.planninginspectorate.gov.uk/wp-content/ipc/uploads/projects/EN010084/EN010084-000606-6.2.10_TEOW_S&N.pdf (accessed 18 February 2020).

MarineCadastre, 2020. Vessel Traffic Data. https://marinecadastre.gov/ais/ (accessed 15 May 2020).

Mazaheri, A., Ylitalo, J., 2010. Comments on Geometrical Modelling of Ship Grounding. 5th Conference on Collision and Grounding of Ships, Espoo, Finland, 14-16 June 2010.

Mazaheri, A., Montewka, J., Kujala, P., 2013. Modelling the risk of ship grounding – a literature review from a risk management perspective. WMU Journal of Maritime Affairs 13(2), 269-297. https://doi.org/10.1007/s13437-013-0056-3.

Mazaheri, A., Montewka, J., Kotilainen, P., Sormunen, O.E., Kujala, P., 2014. Assessing Grounding Frequency using Ship Traffic and Waterway Complexity. Journal of Navigation 68(1), 89-106. https://doi.org/10.1017/S0373463314000502.

Olba, X., Daamen, W., Vellinga, T., Hoogendoorn, S., 2019. Risk Assessment Methodology for Vessel Traffic in Ports by Defining the Nautical Port Risk Index. Journal of Marine Science and Engineering 8, 10. https://doi.org/10.3390/jmse8010010.

Pedersen, P. T., 1995. Collision and Grounding Mechanics. Proceedings of WEMT 95, Copenhagen, Denmark, The Danish Society of Naval Architecture and Marine Engineering.

Pietrzykowski, Z., Uriasz, J., 2009. The Ship Domain – A Criterion of Navigational Safety Assessment in an Open Sea Area. Journal of Navigation 62, 93-108. https://doi.org/10.1017/S0373463308005018.

Qu, X., Meng, Q., Li, S., 2011. Ship collision risk assessment for the Singapore Strait. Accident Analysis and Prevention 43(6), 2030-2036. https://doi.org/10.1016/j.aap.2011.05.022.

Qu, X., Meng, Q., Li, S., 2012. Analyses and Implications of Accidents in Singapore Strait. Journal of Transportation Research Board 2273, 106-111. https://doi.org/10.3141/2273-13.

Rawson, A., Rogers, E., Foster, D., Phillips, D., 2014. Practical Application of Domain Analysis: Port of London Case Study. Journal of Navigation 67, 193-209. https://doi.org/10.1017/S0373463313000684.

Rawson, A., Sabeur, Z., Correndo, G., 2019. Spatial challenges of maritime risk analysis using big data. Soares, C. Guedes (ed.) In Proceedings of the 8th International Conference on Collision and Grounding of Ships and Offshore Structures (ICCGS 2019). vol. 4, CRC Press / Balkema, 275-283.

Rong, H., Teixeira, A., Soares, C., 2019. Risk of ship near collision scenarios off the coast of Portugal. Proceedings of the 29th European Safety and Reliability Conference, Germany, 3660-3666.

Sahr, K.M., White, A.J., 1998. Discrete Global Grid Systems. Weisberg, S. (ed.) In Computing Science and Statistics, 30, 269-278.

Sahr, K.M., White, D., Kimerling, A.J., 2003. Geodesic Discrete Global Grid Systems. Cartography and Geographic Information Science 30(2), 121-134. https://doi.org/10.1559/152304003100011090.

SciPy. 2020. SciPy documentation: scipy.stats.pearsonr. https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.pearsonr.html (accessed 07 August 2020).

Szlapczynski, R., ~~Szlapczynski~~Szlapczynska, J., 2017. Review of ship safety domains: Models and applications. Ocean Engineering 145, 277-289. https://doi.org/10.1016/j.oceaneng.2017.09.020.

Tam, C., Bucknall, R., Greig, A., 2009. Review of Collision Avoidance and Path Planning Methods for Ships in Close Range Encounters. Journal of Navigation 62, 455-476. https://doi.org/10.1017/S0373463308005134.

USCG, 2020. Marine Casualty and Pollution Data for Researchers. https://www.dco.uscg.mil/Our-Organization/Assistant-Commandant-for-Prevention-Policy-CG-5P/Inspections-Compliance-CG-5PC-

/Office-of-Investigations-Casualty-Analysis/Marine-Casualty-and-Pollution-Data-for-Researchers/ (accessed 29 May 2020).

Van Dorp, J.R., Harrald, J.R., Marrick, J.R.W., and Grabowski, M. 2008, VTRA: Technical Appendix D: Expert Judgement Elicitation. https://www2.seas.gwu.edu/~dorpjr/VTRA/FINAL%20REPORT/083108/ VTRA%20REPORT%20-%20Appendix%20D%20083108.pdf (accessed 19 April 2019).

Van Dorp, J., Merrick, J., 2014. VTRA 2010 Final Report. George Washington University.

Wang, N., Meng, X., Xu, Q., Wang, Z., 2009. A Unified Analytical Framework for Ship Domains. Journal of Navigation 62, 643-655. https://doi.org/10.1017/S0373463309990178.

Wang, N., 2010. An Intelligent Spatial Collision Risk Based on the Quaternion Ship Domain. Journal of Navigation 63, 733-749. https://doi.org/10.1017/S0373463310000202.

Xu, Q., Wang, N. 2014. A Survey on Ship Collision Risk Evaluation. Traffic Management Review 26, 475-486. https://doi.org/10.7307/ptt.v26i6.1386.

Yang, D., Wu, L., Wang, S., Jia, H., Li, K., 2019. How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications. Transport Reviews, 39, 755-773. https://doi.org/10.1080/01441647.2019.1649315.

Zhang, W., Goerlandt, F., Kujala, P., Wang, Y., 2016. An advanced method for detecting possible near miss ship collisions from AIS data. Ocean Engineering 124, 141-156. https://doi.org/10.1016/j.oceaneng.2016.07.059.

Zhang, L., Meng, Q., 2019. Probabilistic ship domain with applications to ship collision risk assessment. Ocean Engineering 186, 106130. https://doi.org/10.1016/j.oceaneng.2019.106130.

Zhang, W., Feng, X., Goerlandt, F., Liu, Q., 2020. Towards a Convolutional Neural Network model for classifying regional ship collision risk levels for waterway risk analysis. Reliability Engineering and System Safety 204. https://doi.org/10.1016/j.ress.2020.107127.

***Declaration of Interest Statement**

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

*Credit Author Statement

CRediT Author Statement:

Andrew Rawson: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Validation, Writing – Original Draft, Writing – Review & Editing

Mario Brito: Supervision, Writing – Original Draft, Writing – Review & Editing