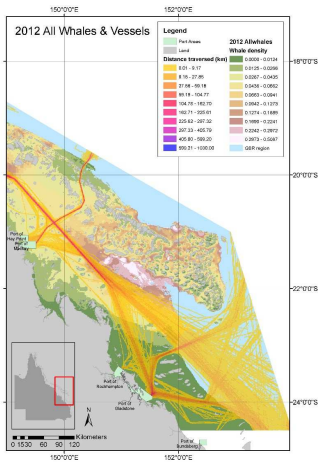
FINAL REPORT

Quantitative assessment of the relative risk of ship strike to humpback whales in the Great Barrier Reef

David Peel[1], Natalie Kelly[1], Josh Smith[2], Simon Childerhouse[3], TJ Moore[4], Jessica Redfern[5]

[1] CSIRO, Hobart, Australia [david.peel@csiro.au; natalie.kelly@csiro.au]
[2] Murdoch University, Perth, Australia [joshua.smith@murdoch.edu.au]
[3] Blue Planet Marine, Nelson, New Zealand [simon.childerhouse@blueplanetmarine.com]
[4] NOAA, La Jolla, California [thomas.j.moore@noaa.gov]
[5] NOAA, La Jolla, California [jessica.redfern@noaa.gov]



Final Report to the Australian Marine Mammal Centre Grants Programme (Project 13/46), Australian Antarctic Division, Australia

10/06/2015

Document Distribution List

Date: 10/06/2015		
Title: Quantitative assessment of the relative risk of ship strike to humpback whales in the Great Barrier Reef		
Company/Organisation	**Position or Location & name of individual**	**Copy**
Australian Marine Mammal Centre	Mike Double, Leader	1
Australian Marine Mammal Centre	Robyn Goyen, Coordinator	2
AMSA, GBRMPA, DoE. IFAW	Humpback whale ship strike risk workshop attendees	3

Document Revision Record

Rev.	Date	Description	Prepared	Reviewed	Approved
1.0	22/05/2015	Draft Final Report for AMMC	DP, NK, JS, SC	DP, NK, JS, SC	SC, JS
1.1	10/06/2015	Pre-Workshop Draft Report	DP, NK, JS, SC, TJM, JR	DP, NK, JS, SC, TJS, JR	SC, JS
1.2	10/07/2015	Final Report	DP, NK, JS, SC, TJM, JR	DP, NK, JS, SC, TJM, JR	SC, JS

Document Reference Number: AMMC-15-FINAL REPORT-Quantitative assessment of the relative risk of ship strike to humpback whales in the GBR-v1.2.docx

Prepared by: David Peel, Natalie Kelly, Josh Smith, Simon Childerhouse, TJ Moore, Jessica Redfern

Last updated: 10/06/2015

Cover picture: Ship struck humpback whale in the Great Barrier Reef. Photo courtesy of and copyright to Dave Paton, Blue Planet Marine.

This document should not be copied or distributed without prior written authorisation from the authors.

Table of Contents

List of Figures

List of Tables

List of Appendices

1. Executive Summary

There is global recognition that ship strike represents a significant risk to some populations of whales around the world. Analysis of ship strike records worldwide demonstrates that humpback whales are the second most frequently reported whale species to be struck by a ship. In Australia, both the east and west coast populations of humpback whales are strongly recovering from commercial whaling during the mid-20th century which resulted in populations nearing extinction. On the east coast of Australia the main breeding ground for humpback whales is within the Great Barrier Reef World Heritage Area (GBRWHA). Both the east and west coast of Australia have also in the past decade experienced considerable coastal and port development associated with an increase in natural resource projects. It is due to substantial coastal development and port expansions related to the mining industry that UNESCO were considering listing the GBRWHA on the 'List of World Heritage in Danger' and are currently monitoring Australia's commitment to its sustainability. Along with considerable port expansion along the GBRWHA coastline to meet increasing global demands for coal and liquefied natural gas (LNG), there is projected to be substantial increases in shipping traffic. Conservative estimates are predicting a doubling of shipping traffic by 2025, albeit not for all Queensland ports. Considering the rapid rate of increase of the east Australian population of humpback whales (approximately 10.9% increase per annum), there is potential for increased interaction between humpback whales and shipping traffic and increased risk of ship strikes to the whales on their breeding ground.

To understand the risk of ship strike to humpbacks in the GBRWHA, it is necessary to understand the distribution and densities for both whales and shipping. This report uses current knowledge on the distribution of humpback whales within the GBRWHA from aerial survey data from 2012 and 2014 and contemporary (2012-2014) shipping traffic data of ships travelling within the GBR to provide estimates of relative risk of ship strike to humpback whales within the GBR. Using the aerial survey data, density surface models were developed to identify whale distribution and density, and then extrapolated to unsurveyed areas for a whale density prediction throughout the whole GBR. The most influential covariates were bathymetry, sea surface temperature and sea surface height anomaly. The models predicted that higher densities of humpbacks were more likely to be found in shallow water (e.g., 20-60 m deep), in waters of a sea surface temperature of 21-23°C and with a sea surface height anomaly of approximately 0.05 m. These models were shown to be consistent with previous spatial habitat models developed (using incidental sighting data) in predicting the distribution of whales. Nevertheless, there were some limitations in the density surface models which resulted in it not being possible to reliably predict for the southern GBR area offshore of Gladstone.

Through a collaboration with the Australian Maritime Safety Authority (AMSA), high quality Automatic Identification system (AIS) vessel tracking data from the Craft Tracking System (CTS) dataset was acquired for the months of July, August and September and the years 2012-2014. Overall, we included over 2.3 million data points post-filtering in the analysis from vessels ≥ 80 m in length[1]. To quantify relative risk, as opposed to absolute risk, we implemented two different metrics comprising an existing risk metric based on the idea of co-occurrence of whales and vessels, and a more complex, probabilistic framework providing a relative index of the expected number of ship strike mortalities. The measure of co-occurrence assumes that when other variables are constant spatially, the degree of overlap between ship and whales should be proportional to risk. Hence, we are considering relative risk rather than absolute. This was undertaken by multiplying the total

[1] Typically, larger vessels pose the highest risk of causing death to humpbacks from ship strike which is consistent with overseas experience. By removing smaller vessels, estimates of risk potentially may be an underestimated; however, if vessel movements for these smaller vessels follow the same general distribution and density of the larger vessels, then this assumption is unlikely to cause a bias in relative risk

distance traversed in a 1km x 1km grid cell by shipping by the number of whales in the cell. The analysis shows that the areas of highest relative risk coincide with offshore areas around the two major ports on the Queensland coast spanning the offshore area between the Whitsundays to south of Mackay, offshore of Shoalwater Bay. One limitation of this risk assessment is that there are no results for the offshore areas to the south of Gladstone due to a lack of whale data. When considering relative risk for groups with calves versus groups without calves, there were no major differences in the spatial distribution between the two groups. As such, once the risk is standardised for the number of total animals of each group type, groups with calves are no more susceptible to ship strike risk than non-calf groups. However, this assumes that our whale density models are correct, and also that both group type are equally likely to be struck (as it does not consider differing risk of ship strike due to differences in whale behaviour).

When considering overall relative risk of ship strike, it was evident that cargo vessels provided the single largest contribution. However, this is not to say individual cargo vessels pose a greater risk than other vessel types, since the co-occurrence measure does not take vessel characteristics into account (i.e., speed, beam, etc.), the difference is a consequence of the large number of cargo vessels relative to tanker and passenger. When we look at risk per km travelled there did not seem to be any discernible difference between individual cargo vessels and tankers based on co-occurrence alone. There was some indication of a higher per km co-occurrence risk for passenger vessels and this would need to be analysed further. In terms of risk it is the overall cumulative risk that is important and so cargo and tanker vessels are of the most concern.

Based on the co-occurrence maps it appears the area of greatest relative risk is two areas located approximately 120km to the North and 120km South of Mackay. After examining the whale habitat models it is clear these correspond to where shipping traverses two higher predicted whale density areas.

As a proof of concept, we also developed and tested a more advanced approach that uses a probabilistic framework to provide a relative index of the number of expected whale fatalities. Overall this approach provides similar results to the broad scale maps using the co-occurrence index, although notably there are some fine scale differences. The advantage of this approach is that this measure considers vessel speed so is a better metric to compare risk from different vessel types and can easily accommodate difference in strike and fatality heterogeneity (e.g., differences between vessel types and/or whale group types). At this point, the index is simply a relative metric across the study area useful to comparing relative risk and cannot and should not be inferred as an estimation of actual mortality. While this step is potentially possible, it would require that considerable data specific to this issue to be collected before the leap between relative and actual mortality rates could be made.

The final relative risk values are the combination of whale and shipping density, and both of these components contain a degree of uncertainty. Quantifying uncertainty is important not only to indicate how much trust should be given to the overall results, but also when it comes to using the application in informing spatial decisions to manage risk. Our approach will capture vessel temporal variability in space much better than variability in whale distribution (as we only had two whale seasons to assess variability). The largest source of uncertainty is likely to be related to inter-annual variation in the spatial distribution of whales as there is little data for which to estimate uncertainty; without additional surveys replicating the coverage of previous ones, this will be difficult to quantify.

2. Introduction

The Great Barrier Reef (GBR) supports significant biodiversity at a global level and is recognised by UNESCO as a World Heritage Area. The marine wildlife values of the Great Barrier Reef World Heritage Area (GBRWHA) are a major reason for the region's World Heritage Listing, which includes internationally significant populations of migrating humpback whales (*Megaptera novaeangliae*) that also breed within this region. In Australian waters, humpback whales are protected within Commonwealth waters under the EPBC Act. Humpback whales on the east Coast of Australia are increasing at a rate of around 11% per annum with numbers predicted to double in approximately 7 years if this increase continues at its present level (Noad et al. 2011). While this is positive news for the whales, it is likely to lead to increased interactions with human activities in the GBRWHA including port development, shipping, and tourism.

In parallel with increasing whales, there has also been a substantial increase in coastal and port development and an associated increase in recreational and commercial shipping along the coastline of the GBRWHA. These projected increases represent a major management issue faced by both the Federal and State governments. A recent analysis of data for 2012/13 showed over 11,000 vessel movements through likely humpback whale habitat within the GBRWHA; an average of 30 movements per day (NESMG 2014). While this represents a significant number of vessel movements, the picture for the future is even more concerning with estimates that vessel movements through the region will almost double by 2020 (GBRMPA 2013).

There is global recognition that ship strike represents a significant risk to some populations of whales around the world with the most well documented example being the North Atlantic right whale with the major cause of population decline directly linked to ship strike (Laist et al. 2001). Analysis of ship strike records worldwide demonstrates that humpback whales are the second most frequently reported whale species to be struck by a ship (Vanderlaan & Taggart 2007). Williams and O'Hara (2009) note that collisions with vessels cause serious injury and mortality in many cetacean species. Quantifying the population level extent of ship strike mortality, however, is notoriously difficult; collisions are frequently unnoticed, and consequently go unreported (Laist et al. 2001; Panigada et al. 2007; Vanderlaan & Taggart 2007). Ship strikes can jeopardise the viability of small populations (Fujiwara & Caswell 2001), and the importance of the topic is reflected in its appearance in the terms of reference of both the Scientific and Conservation Committees of the International Whaling Commission (IWC). The Australian Commonwealth has also recently commissioned the development of National Ship Strike Strategy for large cetaceans to develop options for addressing this issue.

To understand the risk of ship strike to humpbacks in the GBRWHA, it is necessary to understand both distribution and densities for both whales and shipping. Limited data from land and aerial surveys indicate that later in the breeding season (Sept/Oct) whales are closer to the coast than the reef and females with calves are common, exhibiting resting and milling behaviour (Noad et al. 2009). The distribution of humpback whales, specifically mothers with calves, in the southern GBRWHA later in the breeding season remains undetermined and therefore extending distributional information on an increased spatial scale and including temporal coverage of key cow-calf habitat was a priority for this project. However, there have been recent improvements in our understanding of the distribution of humpback whales on their breeding ground in the GBRWHA through the use of spatial habitat models and validation of this from satellite tagged whales and dedicated aerial surveys (Smith et al. 2012). The aim of this research is to determine the distribution of humpback whales at the peak of the breeding season in late July to early August. The main wintering aggregation and high density area of humpback whales in the GBRWHA has been identified in offshore waters in close proximity to coastal areas that are undergoing significant development, including port expansions for coal and liquefied natural gas export.

There are predicted to be increasing levels of shipping associated with the export of natural resources as well as increasing levels of recreational vessel registrations in areas adjacent to the GBRWHA. It is due to substantial coastal development and port expansions related to the mining industry that UNESCO is closely monitoring Australia's commitment to the sustainability of the Great Barrier Reef as a World Heritage Area.

It is uncertain whether these increases in vessel movements may have long-term implications for the eastern Australian humpback whale stock, presently recovering at high rate of increase. With an increase in both shipping and the humpback whale population, there is potential for this to lead to an increase in ship strikes and contribute to cumulative impacts associated with underwater noise and disturbance/displacement from critical habitat. These issues have been identified as important issues for management agencies such as the Great Barrier Reef Marine Park Authority (GBRMPA). Ship strikes involving large vessels and cetaceans may result in death or serious injury to individuals with the level of risk depending on whale density, behaviour, the time of year, vessel density and vessel speed.

While the focus of this project is on assessing the risk of ship strike for humpback whales in the GBR, the major outcome is a flexible modelling framework that can be used to quantitatively assess the risk of ship strike across a range of species and areas, including tools to estimate risk projected into the future to account for increasing vessel traffic and population recovery. Fundamentally, this research builds on and extends two pieces of existing published research on the quantitative assessment of ship strike risk to baleen whales (Redfern et al. 2013 in *Conservation Biology*) and on the quantitative assessment of humpback whale habitat in the GBR (Smith et al. 2012 in *Marine Ecology Progress Series*). We have specifically chosen this approach to build on existing data and proven techniques while extending the application of this research to the entire GBR and increasing the scope to investigate elements such as the incorporation of uncertainty, vessel speed and projected increases in both humpbacks and vessel movements.

2.1 Objectives and outputs

The primary objectives of the project (as stated in the original proposal) are:

- Develop and implement a modelling framework to conduct a quantitative assessment of the relative risk of ship strikes to humpback whales in the GBRWHA using current distribution data from the peak times of the breeding season; and
- To determine the coastal distribution of humpback whales around major coastal and port areas in the GBRWHA to assess temporal changes in whale distribution and assess the relative risk of ship strike in inshore areas.

The outputs of the project (as stated in the original proposal) are:

- Software code for modelling framework for the assessment of relative risk of ship strike for humpback whales in the GBR;
- Technical workshop with NOAA spatial modellers and developers of risk assessment models;
- Report on the estimates of relative risk derived from the modelling framework incorporating projected increases in shipping and whale abundance;
- Report and updated software code documenting extensions to modelling framework to include uncertainty and ship speed;
- Report on density estimates of humpback whales in inshore areas from spatial modelling and comparison with existing distribution data;
- Report on the estimates of relative risk derived from the modelling framework incorporating new data, projected increases in shipping and whale abundance; and
- Report describing recommendations for potentially mitigating ship strike risk for humpback whales in the GBR.

3. Methodology

3.1 Humpback aerial surveys

The Great Barrier Reef World Heritage Area is a large area which has made systematic surveys of the entire area prohibitively costly. Consequently, much of our knowledge of the distribution of humpback whales in the GBRWHA has, until recently, been based on incidental aerial and vessel sightings. A predictive spatial habitat model was developed using opportunistic presence-only whale sighting data from the Coastwatch BPC aerial surveillance program, which through validation by humpback whale satellite tagging data appeared to reliably reflect the whales' distribution (Smith et al. 2012). However, it was difficult to determine whether there was adequate and unbiased sampling of the entire GBRWHA and further validation of the model was required with dedicated, systematic surveys. The aerial survey data presented in this report consists of two separate years of survey data (2012 and 2014) from surveys with different objectives. The survey did not cover the entire area of the GBRWHA but rather sub-sampled within specific regions. The main objective of the 2012 aerial survey was to validate the accuracy of the predictive spatial habitat model. This was undertaken by surveying three main areas that represented areas predicted to have low (Port Douglas), medium (Townsville) and high (Mackay) habitat suitability, which would reflect the density of whales within each area if the habitat model is reliable (Figure 1). The objective of the 2014 aerial survey (which was directly funded as part of this project) was to determine the coastal distribution of humpback whales around major coastal and port areas within a region in the GBRWHA of high whale density to assess temporal changes in whale distribution and allow for the assessment of ship strike risk in inshore areas (Figure 1).

3.1.1 Survey methodology

The aerial survey was performed using a Partenavia Observer P-68B six-seater, twin engine, high-wing aircraft which had bubble windows fitted in the mid seats and flat windows at the rear seats. Surveys were flown in passing mode at a ground speed of 100 knots and a height of 1000 feet in an attempt to improve the ability to identify calves. Transects were spaced 20 km apart and orientated between approximately 22 and 48 degrees from the coastline to survey across the depth gradient, extending offshore from the coastline to the outer reef (Figure 1). Transects within each survey area were undertaken starting with the southernmost transect first, flying north to minimise the probability of double counting animals migrating south through the study area. Humpback whales were the main focus of this survey and consequently the surveys were designed to maximise detection of this species. However, other species of marine megafauna were also counted, including other whales, dolphins, dugongs and sharks.

The survey team consisted of four dedicated observers and a survey leader, constituting a double platform observer configuration. This arrangement allowed the sightings of the two observers on each side of the aircraft to be independent and perception bias to be calculated, whereby observers fail to detect animals even though they are available for detection (Pollock et al. 2006). The two primary observers were seated in the middle seats and the two secondary observers in the rear seats. The survey leader was situated in the front seat next to the pilot and entered all sightings called by the primary observers into a pocket computer using a specialised program developed for humpback whale aerial surveys. The observers and the survey leader communicated via aviation headsets connected to two portable intercoms. Each intercom was connected to a separate track of a two-track digital voice recorder to record the flight audio. During survey mode when 'on effort', the flight leader is in audio contact only with the two primary observers, whereas the secondary observers are acoustically and visually (a black curtain) isolated from the primary observers.

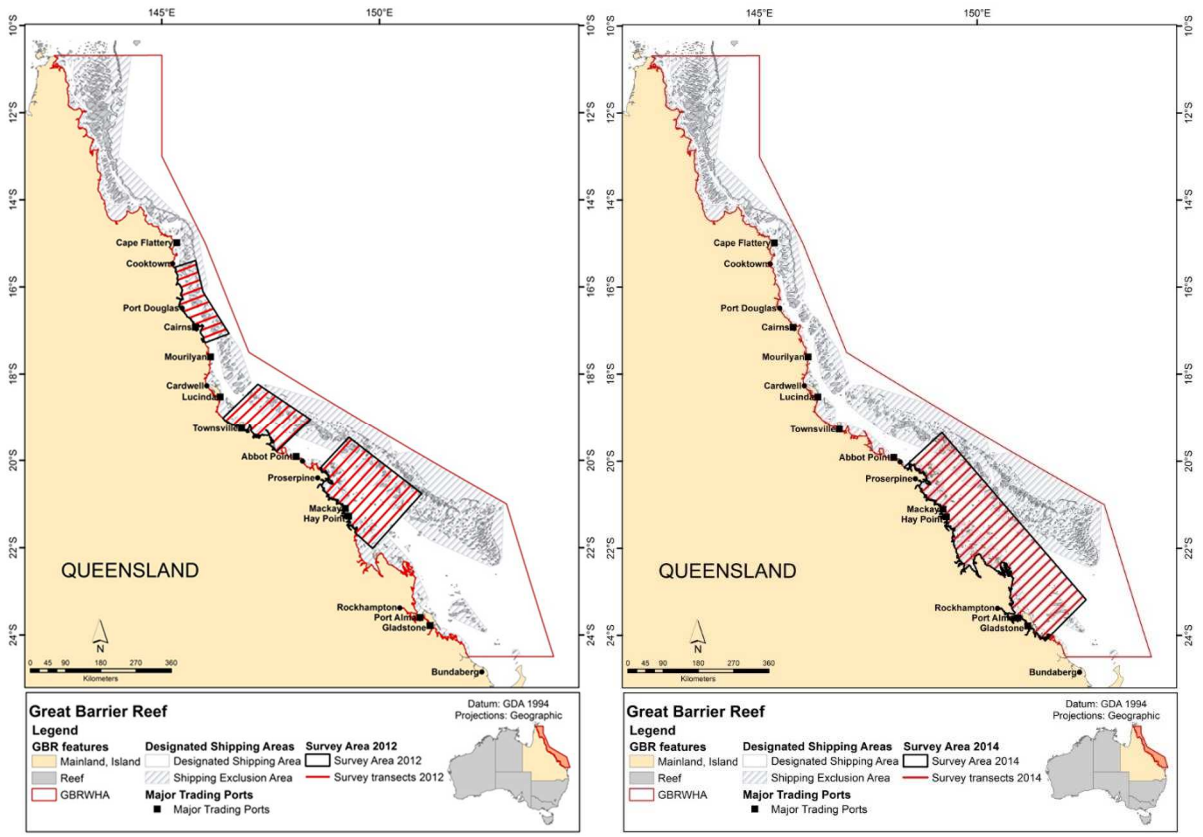


Figure 1: Map of the survey area and transects that were flown during 2012 and 2014 aerial surveys

3.1.2 Analysis of aerial survey data - Habitat suitability/density estimation

The distribution and variations in densities of humpback whales in the GBR were modelled using the 'count method', as described in Hedley & Buckland (2004). The count method is based on two separate statistical models:

- the fitting of a detection function (via distance sampling on perpendicular sighting distances, and other covariates that may influence detectability) to estimate the 'effective strip width' (Marques & Buckland 2003, 2004); and
- the fitting of a 'spatial' model, or density surface model, which involves describing numbers of sightings within small segments of tracklines estimates using generalised additive models (e.g., Wood 2006). These use a smooth over geographical space and/or other potentially informative environmental covariates, and an offset term provided by the effective strip area of each segments[2], informed by the estimate of effective strip widths from the detection function.

Detection probabilities, and corrections for perception bias, were estimated using MRDS models (Mark-Recapture Distance Sampling, as described in Laake & Borchers (2004) and Burt et al. (2014)) using the MRDS package (Laake et al. 2015) in R (R Core Team 2015). As mark-capture (MR) and distance sampling (DS) methods were used during the survey, we were able to assume that the probability of detecting a group that was located on the trackline was less than 1 (i.e., $g(0) \leq 1$). Observers in the front and back positions, on both sides of the aircraft, were isolated from each other for both audio and visual cues, and, as such, are considered independent. The observers in the front positions, on both sides of the aircraft, were treated as the primary observers, and the

[2] Each segment was 10km long

observers in the back positions the secondary observers. Under the assumption there was no responsive movement of animals owing to the noise and movement of the aircraft, point-independence (i.e., between what the primary and secondary observers would see) at the trackline was assumed a priori (but this was subject to testing). Several covariates were available to potentially improve precision and reduce bias in estimate detection probabilities along track: Beaufort sea state, cloud cover (in octas) and whale group size (i.e., multiple-covariate distance sampling, or MCDS, as described in Marques & Buckland (2004)). Furthermore, perpendicular distance to sightings was also tested for the MR component. Improvements in detection function fit with binning of Beaufort sea state and group size values was also tested for. Both half-normal and hazard rate forms of the detection function were tested. Permutations of combinations of all variables, in both of the DS and MR components were tested, and a final (best) detection function, was selected by minimising the Akaike Information Criterion (AIC), and by examining various model diagnostics. To improve detection function fit, perpendicular sighting distances were left-truncated at 0.2 km and right-truncated at 4 km. Sightings with uncertain species information were excluded from the analyses.

For the 'spatial' model, track lines were segmented into lengths of approximately 10 km, and the numbers of groups and total animals (including the presence and number of calves) were summed for each segment. For each segment, the total effective strip area was estimated, via the detection function described above.

A GAM describing abundance over geographical/environmental covariates, with a log-link function, can be generalised as:

$$E(\widehat{N}_i) = \exp[\theta_0 + \sum_k f_k\,(X_{ik})]$$

where *E(Ni)* is the expected number of whales (not sightings in these analyses) in the *i*th segment; θ_0 is the intercept term, f_k are smoothed functions over the explanatory/predictor variables (whether geographical or environmental), X_{ik} is the value of the *k*th explanatory/predictor covariate in the *i*th segment. A log of the effective strip area for each segment was used as offset in the GAM. All sightings that were included in the distance analyses were available to be used in fitting the spatial models.

These counts per segment displayed some evidence of overdispersion under the assumption of a Poisson distribution in the GAM (i.e., there were more zeroes combined with higher values in the counts per segment than might be expected under a Poisson distribution). A Tweedie distribution was assumed to account for the overdispersion (Jørgensen 1987), and owing to the length of the along-track segments, we assumed spatial autocorrelation to be negligible. In assuming a Tweedie distribution, a power parameter must be supplied to the model fitting process; this was inferred from inspection of residual plots during exploratory data analysis. Error from fitting the detection function was propagated through to the spatial model using a method described in Williams et al. (2011) and Miller et al. (2013).

Physiographic variables, such as bathymetry, bathymetric slope, and great-circle (geodesic) distances to both the nearest coastline and reef features were estimated for the midpoints of each along-track segment. Monthly mean values of dynamic environmental/habitat predictor variables (remotely sensed) were interpolated to the midpoint of each along-track segment. Daily sea surface temperature (IMOS, 2015a; in °C, gridded at 0.02°), sea surface height anomaly (IMOS, 2015b; in metres, gridded at 0.58°×0.51°), and sea surface chlorophyll a (IMOS, 2015c; mg m-3, gridded at 0.01°) values for the GBR region were averaged at each grid point, across each day of August 2012 and September 2014, to yield month-wise estimates of these environmental covariates. Values of each environmental covariate were converted into rasters in ArcMap 10.1 (ESRI), and matched to the midpoints of each along-track segments.

Because the aim was to extrapolate whale densities throughout GBR, and well beyond the survey regions, inclusion of geographic coordinates as predictor variables was deemed inappropriate. However, fitting purely spatial models (i.e., only based on geographic coordinates, either latitude/longitude or some projected coordinate space) is useful to explore how densities might vary within the survey area at the time of survey. The shapes of any subsequent density surfaces based on environmental covariates should broadly mirror the purely spatial model. Therefore, purely spatial models, based on a projected coordinate space (Albers equal area) were produced for this purpose.

Collinearity in the explanatory variables can lead to spurious parameter estimates, which is problematic if there is a desire to use such estimates for explaining ecological or biological processes. Collinearity in the various spatial/environmental covariates were assessed using multi-panel scatterplots and Pearson correlation coefficients.

Physiographic and environmental variable selection for the GAMs was based on changes in deviance explained with addition/removal of covariates (as fitting method was REML, use of AIC for model selection is not appropriate). This process commenced by fitting a GAM that included all covariates as individual thin-plate regression splines, with null space penalties; any resultant effective degrees of freedom which tend to zero can be dropped from the model (i.e., they are not useful). Combinations of remaining covariates, in individual thin-plate regression splines, and/or interactions, via either thin-plate (when covariates are isotropic) or tensor product smooths, can be tested. Standard model diagnostics were performed on the GAMs, and further model robustness was assessed by predicting for along-track abundance, and comparing to observed whale counts.

As an aim of this project is to estimate the relative risk of ship strike upon a population of animals on their breeding groups, analyses of distribution of inferred densities was also extended to groups that contained at least one calf.

For predictions of whale densities across the GBR, a 1×1 km grid was produced, with each grid point populated with environmental covariate values - as above, for track segments, values of each environmental covariate was matched to grid point in ArcMap.

3.2 Shipping data

3.2.1 GBR shipping routes

Within the north-east region there are three major shipping routes; the Torres Strait Route, Inner Route and Outer Route. The Torres Strait route is the most direct route from south Asia and India to eastern Australia and links into the northern portion of the GBR. The Inner Route runs parallel to the Queensland coast and lies between Cape York in the north and Gladstone in the south. The Outer Route begins at the eastern limit of the Torres Strait (the Great North East Channel), continues southwards through the Coral Sea and re-joins the Queensland coast near Sandy Cape (south of Gladstone). A two-way shipping route in Queensland's Great Barrier Reef (GBR) and Torres Strait has recently been formalised by the International Maritime Organization (IMO) and in effect as of December 2014. The IMO-adopted two-way route extends from the western end of Torres Strait, through the Prince of Wales Channel, the Great Barrier Reef Inner Route and terminates at the southern boundary of the Great Barrier Reef Marine Park. The two-way route is unchanged from that previously chartered in the Torres Strait and northern GBR. A new section of the two-way route has been introduced in the southern portion of the Great Barrier Reef that follows existing traffic patterns. The two-way route formalises the routes that have been in existence since the 1980s, although now provides well defined lanes.

The majority of ships enter and leave the Torres Strait and GBR ports via six main passages:

- Great North East Channel (Torres Strait);

- Grafton Passage (near Cairns);
- Palm Passage (near Townsville);
- Hydrographers Passage (near Mackay);
- Capricorn Channel (near Gladstone); and
- Curtis Channels (near Gladstone).

3.2.2 AMSA AIS data

AIS data was obtained from AMSA in the form of their craft tracking system (CTS) product (See the metadata in Appendix 6). The CTS provides cleaned, processed data[3] sampled to a 5 minute polling frequency. This sample rate seems a good compromise between data set size and spatial uncertainty due to unknown path/locations between polling[4].

A higher quality data series began in June 2012 and so it was decided to start analysis at this point, providing 3 years of data covering 2012, 2013 and 2014. This period also corresponds to the systematic aerial surveys for humpbacks undertaken in the GBRWHA in 2012 and 2014. We restricted shipping data to the 3 winter months corresponding to the Australian humpback whale breeding season (i.e., July, August, and September) although it is known that whales are also present in lower numbers outside of these times.

AMSA provided the data in compressed CSV text files.

Spatially, we acquired the data for the bounding box of the GBR and then clipped it to the GBRWHA management area (Figure 2).

Distance-sampled vs Time-sampled

The AIS data is time based, that is the data sampling is time-based, whereas the metrics of concern in ship strike risk are distance-traversed based. So, to use the AIS data for our application, we need to convert or weight the data to distance rather than time. This is further discussed in section 3.2.6.

[3] In the raw data the AIS system can produce multiple entries for a single location from various satellites etc.

[4] Given a typical average/mean vessel speed could travel at 12 knots, the distance traversed in 5 mins would equate to 1.852 km

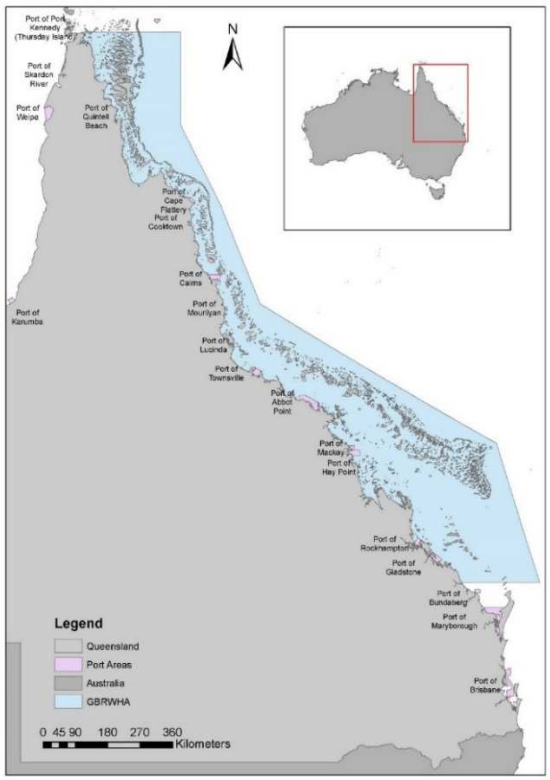


Figure 2: Map of the data bounding box (blue) and GBR management area

3.2.3 Data validation

Summary tables were compiled listing each unique vessel (based on MMSI) and summarising the values it had in the data for various information (e.g., length, beam, draught, type, class, name, IMO, etc.). By doing this we could easily discern vessels with missing or multiple values. For example vessel length was missing in a number of cases. Similarly, we found some issues with shipping type.

Therefore, for Class A vessels in the summary table with incomplete data, we manually looked up the vessels on websites to confirm details (e.g., www.marinetraffic.com/en/ais/index/ships/all , www.fleetmon.com/en/vessels). There were also a few instances of vessels having the same MMSI number, so generally we found it better to search on the IMO number.

3.2.4 Data filtering/Cleaning

The AIS data was supplied by AMSA in a csv file and we then imported, cleaned and filtered it by a number of criteria based on the previous experience of NOAA in analysing shipping data for ship strike risk (see Table 1 and Table 2). This was done in PYTHON/ArcGIS 10.1 libraries using modifications of scripts provided by NOAA (TJ Moore and J Redfern).

Table 1: AIS data filtering criteria used

Filtering	Criteria	Comment
Class	Class A vessels only	Class A are required to use the AIS system and covers the vessels that can potentially be of concern for ship strike and large whales

Vessel Type	Keep cargo, tanker or passenger only	These are the commercial vessels using defined shipping routes travelling 'fast' and most likely to be a ship strike risk to large whales
Vessel Length	≥ 80 m	Vessels of smaller size will be less of a concern for large whales
Speed	> 0.4 knots	The AIS data we have does not have navigational status which can be used to filter out vessels which are not underway (e.g., anchored). Since obviously stationary vessels are of no concern we attempted to remove these with this criteria[5]
Valid MMSI only	201,000,000 ≤ MMSI ≤ 775,999,999	MMSI outside this range are invalid and produced by corrupt data.

Table 2: AIS data error codes identified and set to null.

UnknownCOG Flag	>= 360	To flag unavailable (360) or erroneous (> 360) data for the course over ground field
UnknownSOG flag	1022 or 1023	To flag excessive (>= 102.2 knots) or unavailable speeds in the speed over ground field[6].

3.2.5 Data processing

The processing overview is shown in Figure 3, the points are joined to form lines, and then these are divided over a grid and information summarised. The first step of joining the points is done using python/ArcGIS script (provided by TJ Moore, NOAA). Several criteria are used to remove lines that will be too uncertain, or have data issues (see Table 3).

[5] 0.4 was used to remove stationary vessels that drift. This was not completely successful, but since our ship strike risk metrics all involve distance traversed and/or speed this should not cause any bias. We will discuss this in the results.

[6] Note: This is in the raw AIS data where SOG is recorded as knots (reported to the tenth of a knot)*10. In the data we received, the SOG had been converted to actual knots. So the unknown code was 102.2 and 102.3

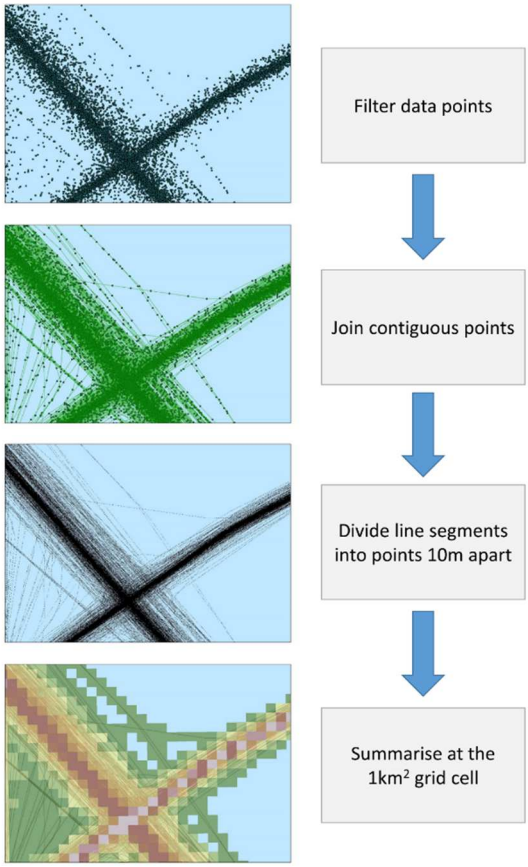


Figure 3: Demonstration of the AIS data processing

3.2.6 Estimating shipping traffic density

To estimate shipping traffic density we formed a 1 km^2 grid over the entire GBR region. This size was chosen based on typical width of shipping lanes, so as to provide enough spatial resolution to distinguish specific lanes. Then we summarise the line data based on how much of each line segment is in each grid cell. Doing this converts, or weights, the data in each cell by distance traversed rather than time (see section 3.2.2)

ArcGIS provides tools to calculate the intersection of lines and polygons, and we have custom written python scripts from NOAA (Moore) to undertake this task. In this application, we typically have around 900,000 line segments and around 360,000 grid polygons for each season and due to only having 32 bit drivers installed, python/ArcGIS had memory issues. Instead we developed some R code to process the shapefiles and produce grid cell summaries and save as a shapefile.

The R code used a numerical approach rather than analytic, by dividing each line segment into points evenly spaced (10 metres) apart, and then simply estimating statistics for all the points contained in each grid cell. For speed and simplicity, all calculations were done in equal-area projected space. This does introduce some bias since distance properties are not maintained in the equal area projection. Estimating the distance distortion across all line segments we found the typical error to be ~0.02% with 99.9% of the data having an error of <0.01%, which we found acceptable given the spatial scale the final risk maps will be queried.

Table 3: AIS line creation validity criteria

Filtering	Criteria	Comment
Max polling interval time (Δt)	If Δt ≤ 30 mins then keep	Although the data is sampled at a poll every 5 mins, due to technical issues on some occasions polling is less frequent. We added this limit as beyond that the path/track of the vessel between the poll locations is highly uncertain
Longer polling time (Δt) but straight travel ΔCOG	If 30 mins ≤ Δt ≤ 60 mins **and** ΔCOG ≤ 5° Then keep	If the polling interval Δt is longer but the vessel seems to be travelling reasonably straight (based on the change in course over ground ΔCOG), we are still reasonably confident we can interpolate where the vessel was between polls
Longer polling time (Δt) but <u>not</u> straight travel ΔCOG	If 30 mins ≤ Δt ≤ 60 mins **and** ΔCOG > 5° Then remove	
Long polling time	If Δt > 60 remove	If the polling interval Δt is too great, we cannot be certain the path the vessel took and so we delete the transect (the code has the option to leave the start and end points, in the data as they are certain locations, we did not use this option for our analysis)
Backwards timetravel and zero length lines	Remove any Δt≤0 or line length = 0	This should not occur but if there is a negative change in time obviously something is wrong and this data is ignored
Ship tracks with apparent positional errors	Distance traversed equates to travel that equates to ≥ 60 knots	Occasionally due to corrupt data, bad GPS fix, or a mix up in reported MMSI from another vessel, vessels can jump at impossible speeds. These are removed.
Land	Leave in the data	Due to GPS errors or corrupt data a very small number of locations correspond to land. Since our grid data used later is for 1 km^2 cells containing ocean, any obvious land points will be filtered out automatically at that stage.

3.2.7 Other potential information/outputs

With further analysis, other information could be extracted that may be of use in a general ship strike context including:

- Time of day;
- Changes in density over time (both intra and inter season); and
- Vessel speed.

3.2.8 Projected increases in shipping traffic

Predicting future shipping traffic requires an understanding and consideration of historical and contemporary market, economic and industry factors and uncertain global economic influences is often what creates uncertainty in predicted shipping volumes. Consequently, the most prudent approach when forecasting shipping traffic is to undertake periodic re-evaluations.

In the north-east region around the GBR, there is a predicted growth trend in shipping activity predominantly as a consequence of increased export trade, with traffic growth highly correlated to commodity growth (Braemar Seascope 2013, PGM 2012). Most projected forecasts have shipping traffic doubling by 2020, although this is not for all ports with varying increases dependent on the port (PGM 2012). Specifically, Abbott Point, Hay Point, Port Alma and Gladstone ports demonstrate the most substantial increases. In relation to the types of commodities that ships are servicing within the GBR, the Queensland commodity market is and will be dominated by coal. In 2015, coal represents 82% of total trade and is predicted to still remain substantial in 2025 at 81% (Braemar Seascope 2013). Consequently, projected shipping traffic volumes predominantly focus on current and future port capability for export and global demand for coal. There is predicted to be an 83% increase in coal exports between 2011- 2025, corresponding with a 58% increase in projected shipping levels (AMSA 2014). While the coal exports contribute to the majority of shipping traffic, a significant increase in shipping is also predicted to occur related to Liquefied Natural Gas (LNG) exports. This is due to significant increases in production capacity, such that by 2020 Australia is likely to be the world's second largest exporter of LNG after Qatar. Consequently, Australian LNG exports are likely to triple over the next five years with LNG processing plants for exports coming online in Gladstone and Curtis Island in 2014-15 (AMSA 2014). Historically, the Port of Gladstone has had no experience with LNG shipping, although the first LNG carriers were introduced in 2014 for Queensland Curtis LNG projects. It is predicted actual output will grow by around 250 per cent up until 2018 with growth in LNG traffic peaking at 500 ships per year in 2020 (Braemar Seascope 2013).

3.3 Other data

A description of environmental data used in the analysis is described in section 3.1.2.

3.4 Statistical analysis

3.4.1 Risk modelling framework

To quantify relative risk we need a suitable metric. We implemented two different metrics as defined in the following sections.

3.4.1.1 Metric 1 – Index of co-occurrence

The first metric implemented was a simple measure of co-occurrence, which assuming other variables are constant spatially, should be proportional to risk. The measure of co-occurrence in a particular grid cell (*i,j*), was simply taken to be the distance traversed in a cell by ships D_{ij}, multiplied by the number of whales in the cell W_{ij},

$$\mathrm{Risk}_{i,j} = D_{i,j} \times W_{i,j}$$

It has been found that in some applications there can be issues with scaling between the shipping and whale components which can lead to the co-occurrence measure being dominated by the shipping data. Similarly, there may be some repercussions from the often skewed nature of the distribution of the shipping term, D_{ij}. There are several approaches that can be used to address this, for example, transforming the data ($D_{i,j}$ and $W_{i,j}$) or categorising them and multiplying the categories. Hence, there are a number of ways to characterise of co-occurrence risk and these alternatives should be investigated to aid in generating a general picture of risk.

3.4.1.2 Metric 2 – Relative index of the expected number of fatalities

Due to unknown parameters and mechanisms (i.e., potential whale avoidance of vessels, dive behaviour, etc.) achieving an absolute probability of fatal ship strike is difficult in this application. Therefore, we propose formulating a relative probability of a fatal strike occurring and from this a measure proportional to the expected number of fatalities (i.e., a relative index). That is, we aim to estimate a probability that is proportional to the true expected probability of fatal ship strike. This allows us to ignore terms/aspects that are unknown but reasonably constant across cells, while still allowing a relative comparison of risk between spatial locations and other comparisons.

Let us consider a single grid cell, in general terms we can think for a given whale, *w,* the probability of a fatal strike from a single vessel, *v,* as the probability of a fatality given there was a strike multiplied by the probability there was a strike, using a conditional probability rule

$$\Pr\left(Fatality_{w,v}\right) = Pr\left(Fatality_{w,v} \middle| Strike_{w,v}\right) \times Pr\left(Strike_{w,v}\right)$$

For the first term, there is some information available in the literature of the probability of fatality from a strike given vessel speed. Both Conn & Silber (2013) and Vanderlaan & Taggart (2007) provide models for $Pr\left(Fatality_{w,v} \middle| Speed_v, Strike_{w,v}\right)$ for large whales.

Looking at the second term, for a strike to occur a whale must be in the same approximate horizontal position (*xy*) as the vessel some time during the period of time the vessel is in the grid cell. In addition, it must be within vertical proximity to the vessel within the water column, which makes it susceptible to a strike based on co-location and the hydrodynamic effects present during such an interaction. Finally, the whale must not avoid the vessel. So we can express the probability of a collision or strike as,

$$\Pr(Strike) = \Pr(\text{Depth}_w \leq \text{Draft}_v) \times \Pr(xy_w = xy_v) \times (1 - \Pr(Avoidance_w | Speed_v))$$

Since little is known about many these factors (e.g., Dive rate and hence when the animals are near enough to the surface to be struck, whale behaviour/reaction to vessels etc.) this probability will be difficult to quantify. However, we can obtain an approximation that is in theory proportional to the real probability and therefore we can derive a relative overall probability that will be useful for the comparison of relative risk between spatial areas.

Probability of surface availability $\boldsymbol{Pr(Depth_w \leq Draft_v)}$

For the probability that the whale is at a depth where a collision could occur, we propose to use a simple multiplier $\Pr\left(\text{w}_{Depth} \leq \text{v}_{Draft}\right)$ in a similar fashion to in the literature (e.g., van der Hoop et al. 2012). However, what "surface" means in the context of ship strike risk assessment is unclear, considering recent work on the hydrodynamics of ship strike (Silber et al. 2010).

With dive depth profile data from tagging, we could actually derive a function relating $\Pr\left(\text{w}_{Depth} \leq \text{v}_{Draft}\right)$ to v_{Draft} of individual vessels. Values in the literature for time near 'surface' range between 0.66-0.71 for humpback whales on non-feeding grounds (e.g., Andriolo et al. 2006; Baird et al. 2000). However, based on the hydrodynamics work of Silber et al. 2010 the vertical zone could be 3.3 * draft which would increase the probability, based on rough calculations, to approximately 0.85.

However, in our application, the aerial survey itself adjusts for surface availability. That is the whale model we have provides relative abundance as it does not correct for animals that were not at the surface to be observed. Based on some very rough approximations (see Appendix 5.1 for further discussion), the adjustment will be of the order 0.90-0.95. However, more work is needed to better quantify this adjustment.

Probability of whale avoidance $\boldsymbol{Pr(Avoidance_w|Speed_v)}$

Since little is known about potential humpback whale vessel avoidance, we shall assume whales do not avoid vessels, and therefore $\Pr(w_{avoidance}|v_{Speed}) = 0$. There is some evidence in the literature suggesting that both southern right whales (Nowacek et al. 2004, Vanderlaan and Taggart 2007) and blue whales (McKenna et al. 2015) show very little avoidance. Even if humpbacks do have avoidance behaviour and if it is similar spatially, and there is no large spatial difference in vessel speeds of the grid cells being compared, then this assumption will not cause bias. However, it is likely that there is spatial differences in whale behaviour but the magnitude is unknown (i.e., mothers and calves in shallow water could show different avoidance than males in deeper waters). However, any possible bias may be mitigated in the final probability of fatality as the $Pr(Fatality_{w,v}|Strike_{w,v})$ is small for slow vessels where potentially whale avoidance may come into play but will be irrelevant for faster vessels where fatality is assumed to be 100%.

Probability of whale and vessel being in the same place $\boldsymbol{Pr(xy_w = xy_v)}$

We now look at the probability that a vessel and a whale will occur in the same place at the same time, $Pr(xy_w = xy_v)$, disregarding depth. One approach to estimate collision probability is to use a simulation based approach (Van der Hoop 2012). We feel for our purposes any simulation would only provide an approximate relative probability due to the simplification and the complexity of the real whale-vessel interaction. Therefore, since we are aiming for a relative probability anyway, we see no loss at this stage from using an abstract analytic approximation rather than a simulation.

So let us assume random distribution of animals within the grid cell, given the size of the grid cells (1 km^2) this seems a reasonable assumption. Furthermore, we assume that the whale is stationary. Based on typical humpback (i.e., Noad & Cato 2007 found Queensland migratory humpback swim speeds of 1.35 knots for non-singers, 2.16 knots for singing animals) this simplification seems reasonable. Furthermore if whale movement is random, the stationary assumption should not introduce bias in the expected number of collisions. Given these assumptions/simplifications the probability the whale and vessel come to occupy the same place can be expressed in terms of swept area relative to the area of the grid cell, Area$_g$,

$$\Pr(xy_w = xy_v) \propto \frac{Swept\_Area_v}{Area_g}$$

$$= \frac{Beam_v \times Distance_v}{Area_g}$$

This measure could be further refined (e.g., van der Hoop (2012) and others incorporate whale dimensions and given the work by Silber et al. (2010), some measure of vessel volume would seem appropriate). While our initial approximation represents a large simplification, as long as the quantity is approximately proportional to strike probability and any bias is consistent across grid cells, we can still obtain a relative probability.

So in summary, for a single whale *w* and vessel *v* in a single grid cell

$$\Pr(Fatality_{w,v}) \propto \frac{Beam_v \times Distance_v}{Area_g} \times \Pr(Depth_w \leq Draft_v) \times Pr(Fatality_{w,v}|Strike_{w,v})$$

If we now consider *T* vessels in the grid cell then we want to estimate,

$$\Pr(Fatality_w) = \Pr\left(Fatality_{w,v=1}\right) \; or \; \Pr\left(Fatality_{w,v}\right) \; or \; \ldots or \; \Pr\left(Fatality_{w,v=T}\right)$$

By the probability rule for 'or' (see Appendix 5.2)

$$\Pr(Fatality_w) = 1 - \prod_{v=1}^{T}\left[1 - Pr\left(Fatality_{w,v}\right)\right]$$

Now this gives the relative probability of a single whale in the grid cell being fatally struck. So to estimate the expected number of fatalities for the grid cell for *w* whales, we consider this a Binomial with probability=$\Pr(Fatality_w)$ and size = *w* and the expected value is given by

$$\mathbb{E}(No.Fatalities) = w \times \Pr(Fatality_w)$$

We could also work out, given *w* whales in the grid, what the relative probability of a strike occurring is. See Appendix 5.3 for further details.

3.4.2 Relationship between metrics

To compare the co-occurrence metric to the expected number of fatalities, looking at the formulation of the Expected number of fatalities we can see that

$$\begin{aligned}\mathbb{E}(No.Fatalities) &= w\left(1 - \prod_{v=1}^{T}\left[1 - Pr\left(Fatality_{w,v}\right)\right]\right)\\ &= w\left[\sum_{v=1}^{T} Pr\left(Fatality_{w,v}\right) - \Phi(.)\right]\end{aligned}$$

where $\Phi(.)$ is simply all the terms that adjust for the intersection of events (e.g., the $\Pr(A \cap B)$ terms in Appendix 5.2). Now if we assume all the terms that go into $\Pr\left(Fatality_{w,v}\right)$ are constant (i.e., all vessels same beam and speed, etc.) then $\Pr\left(Fatality_{w,v}\right)$ will be the same for all *w* and *v*, lets denote this by λ and that means Φ will only depend on λ and *T*.

$$\begin{aligned}&= w\left[\sum_{v=1}^{T} \lambda - \Phi(\lambda, T)\right]\\ &= w[T\lambda - \Phi(\lambda, T)]\\ &= wT\lambda - w\Phi(\lambda, T)\end{aligned}$$

And so a relative measure is given by (dividing by λ)

$$\mathbb{E}(No.Fatalities) \propto wT - \frac{w\Phi(\lambda, T)}{\lambda}$$

The co-occurrence metric is simply $w \times T$ so we can see that the repercussion of using a co-occurrence metric has some potential bias, due to the term $-\frac{w\Phi(\lambda,T)}{\lambda}$ being ignored. This corresponds to the assumption that the risk is linearly proportional to the density of ship traffic and the number of whales, which as can be seen here that is not strictly true since as *w* and *T* increase the risk will start to asymptote. Based on calculations for a given $\Pr\left(Fatality_{w,v}\right)$ and *T* it would appear for small $\Pr\left(Fatality_{w,v}\right)$ the bias only becomes an issue for very large *T* and therefore in this application co-occurrence should be a reasonable proxy for risk.

3.4.2.1 Overall relative risk calculation

To calculate co-occurrence risk index or relative index of the expected number of fatalities we match the whale and shipping densities for each 1 km^2 grid cell for a specific year (season). Since our aim is to estimate general ship strike risk, not necessarily historical risk for specific years, there is no need

to match whale survey years with vessel data years. Instead it is possible to get a general indication of risk for every combination of ship data year (winter 2012, 2013, and 2014) and whale year (winter 2012 and 2014) which we evaluated. These fine-scale results were then also summarised at both 50x50 km grid cells and 150x150 km square bins and also cumulative overall totals of risk were calculated for subsets of the data (e.g., vessel types and whale group types).

3.4.2.2 Mapping and assessing relative risk

We can map the risk measures in sections 3.4.1.1 and 3.4.1.2 using a fine spatial resolution (e.g., across the GBR using a 1 km^2 grid cell). However, these maps do not suit broad scale comparisons and were better for comparing/investigating small spatial areas. This is because the fine-scale maps show a comparison between each small grid cells rather than give an indication of the total relative risk for an area or a lane. In a statistical sense the maps are good at indicating the peaks and troughs of the spatial risk distribution but to make general comparisons the risk needs to be integrated over the area or subset of the data being compared.

To better explain this, consider crossing a road, you have a choice to cross where it is a single lane and there is a 20% chance of being run-over or cross in another place where there are 8 lanes each with a 5% probability of being run-over. If we were to look at a risk map with a resolution of a lane width, the single lane part of the road would stand out as high risk whereas the eight lane road would appear lower risk. However, in fact the cumulative risk is much higher for the 8 lane road since you have 8 chances of being run-over so the total is 34%[7].

So the procedure for investigating risk followed these steps:

1. Calculate relative risk totals and look at general comparisons between years and various subsets of the data;
2. Examine broad scale maps to identify on a broad-scale where the cumulative co-occurrence risk was higher; and
3. Examine fine-scale maps to identify parts of the lanes/locations within the higher risk locations identified in step 2 that are causing the higher risk.

3.4.3 Projected future relative risk

Predicting future relative risk based on projected growth rates can be difficult, given we do not know exactly how either of these increases will play out spatially. For example, will whales spread out and use more areas as their numbers increase or simply increase in density within the areas currently being utilised. There is some information on predicted increases per port (e.g., Braemar Seascope 2013) and this could be used in calculating future risk by simply breaking our risk calculation into regions and increasing each region based on predicted increases and show changes in fine scale relative risk maps and calculate total GBR-wide changes in risk.

For this report we chose to make the broad assumption that there are no changes to spatial distribution with increases in whale numbers and ship traffic. Similarly, for this initial calculation we assume increases are spread across all vessel types and lengths or calf/non-calf proportions.

Given this assumption and using the relative co-occurrence measure the calculations are simple as we can ignore spatial aspects and simply look at the total co-occurrence risk. As per section 3.4.1.1the measure of co-occurrence in a particular grid cell (i,j), is given by the distance traversed in a cell by ships D_{ij}, multiplied by the number of whales in the cell W_{ij},

$$\mathrm{Risk}_{i,j} = D_{i,j} \times W_{i,j}$$

So now if we add a year component *t*,

[7] Based on pr(hit at some point)=1-probability(not being hit), the probability of not being hit = (1-0.05)^8

$$\text{Risk}_{i,j,t} = D_{i,j,t} \times W_{i,j,t}$$

Now given an annual proportional increase of r_D and r_W for ship traffic and whale abundance respectively, then assuming the present is year t=0,

$$D_{i,j,t} = D_{i,j,0} \times (1 + r_D)^t$$

and similarly

$$W_{i,j,t} = W_{i,j,0} \times (1 + r_W)^t$$

So

$$\begin{aligned}\text{Risk}_{i,j,t} &= D_{i,j,0} \times (1 + r_D)^t \times W_{i,j,0} \times (1 + r_W)^t \\ &= D_{i,j,0} W_{i,j,0} \times (1 + r_D)^t (1 + r_W)^t\end{aligned}$$

This is the increase in each grid cell, if we consider the GBR as being made up of G equal sized grid cells, then total co-occurrence risk is given by

$$\begin{aligned}\text{Risk}_t &= \sum_{\forall i.j} \text{Risk}_{i,j,t} \\ &= \sum_{\forall i.j} \left[D_{i,j,0} W_{i,j,0} \times (1 + r_D)^t (1 + r_W)^t \right] \\ &= (1 + r_D)^t (1 + r_W)^t \sum_{\forall i.j} \left[D_{i,j,0} W_{i,j,0} \right] \\ &= (1 + r_D)^t (1 + r_W)^t Risk_0\end{aligned}$$

Therefore, the multiplicative increase in the co-occurrence risk at any given year t compared to the present is

$$(1 + r_D)^t (1 + r_W)^t$$

3.4.4 Uncertainty

To provide an indication of the uncertainty, we looked at the relative risk evaluated for each combination of shipping data year (winter 2012, 2013, and 2014) and whale model year (winter 2012 and 2014)[8]. Then for each 1 km^2 cell we report the mean, minimum, and maximum risk observed. Also we calculate the range of the estimates standardised by their mean.

3.4.5 Assumptions

Assumptions are discussed in detail in the Results, section 4.5.6.

4. Results

4.1 Technical Workshop

A 4 day Technical Workshop was held in Hobart 18-21 November 2014 to discuss the project, work together in the development of analytical approaches and to explore potential extensions to the work. Three of the four primary researchers were present with the fourth joining by teleconference.

[8] We could also add whale model uncertainty at this stage by either adding noise to the whale densities based on the model prediction standard errors. Due to time constraints this was not completed at this stage.

In addition, our two USA based collaborators Jessica Redfern and TJ Moore also joined the workshop. Overall, it was an excellent success and was a very productive time. A further Technical Workshop was held in June 15-19 2015 to discuss the results with Stakeholders and interested researchers with a view to exploring future extensions to the project.

4.2 Humpback aerial surveys

4.2.1 General survey results

The 2012 aerial survey was undertaken in the GBRWHA offshore of Mackay, Townsville and Port Douglas over 8 days during 3 to 10 August. The total areal coverage for each survey area was: Mackay (34,626 km^2), Townsville (17,126 km^2) and Port Douglas (11,971 km^2). In total, there were 575 sightings of whale groups by front and rear observers (includes resight data).

The 2014 aerial survey was undertaken offshore of Gladstone and Mackay over 11 days from 26 August to 5 September. The survey was undertaken later in the breeding season in an attempt to determine any coastal dependence by whales as the season progressed. The total areal coverage for this survey area was 72,752 km^2. In total, there were 417 sightings of whale groups by front and rear observers (includes resight data). The amount of flying time and percentage of 'on effort' time spent surveying within different sea states is presented in Table 4.

Table 4: Summary data for humpback aerial surveys in 2012 and 2014

Survey year	Flight time (hrs)	On effort flight time (hrs)	Beaufort sea state (percentage of on effort time)					
			0	1	2	3	4	5
2012	27.5	15.75	0.0	68.5	22.6	6.0	2.8	0.1
2014	32	18.3	1.4	77.2	9.5	10.7	1.1	0.0

From the combined aerial surveys there were a total of 637 sightings of humpback whale groups, 365 sightings (589 individuals) in 2012 and 272 sightings (461 individuals) in 2014. The distribution of sightings are given in Figure 8. There were a total of 159 sightings of humpback whale groups with at least one calf present, 100 sightings (121 individuals) in 2012 and 59 sightings (218 individuals) in 2014. The corresponding encounter rates are given in Table 5. The encounter rates for all sightings were slightly lower in 2014, as compared to 2012. Under the assumption that encounter rate scales reasonably linearly with group densities across survey years, there is no significant difference (at the 0.05 level) in the overall density of groups between 2012 and 2014, nor for densities of groups containing calves (using a two-sample *Z* test).

Table 5: Encounter rates for all sightings, and for sightings with a calf present, for both surveys. Standard errors estimated at the transect level

Survey	Total survey length (km)	Mean encounter rate (SE): all sightings	Mean encounter rate (SE): sightings with calves
2012	3037.4	0.1429 (0.0031)	0.0329 (0.0009)
2014	3612.3	0.1224 (0.0037)	0.0163 (0.0007)

4.2.1.1 MRDS and the fitting of detection functions

Mark Recapture Distance Sampling (MRDS) approaches were used to estimate humpback density. Both half-normal and hazard rate key functions were tested. Group size was considered in two different ways: (1) as both a raw count and (2) as a re-classified value based on ground sizes of 1, 2 and 3+ animals (NB. for both size classifications, number of animals includes the presence of any

calves). Other covariates considered included Beaufort sea state (both as a raw level and as a reclassified value of 0-1 (good-excellent sighting conditions), 2+ (fair to progressively worse)), cloud cover and survey. Survey was also included as a categorical variable to test for whether the observer team (which varied between surveys) had an influence on the scale properties of the detection function. The best fitting MRDS model was selected using the AIC (i.e., the lowest).

The MRDS modelling process was repeated for both point and full independence assumption (well, repeated in the context that it used the same R code – however, the 'DS' component is not used under the assumption of full independence and the full independence assumption was found to be routinely violated. Therefore, the point independence assumption was subsequently used in all runs.

The fitting of a detection function was based on sighting data pooled across both survey years. Of the 637 sightings summarised in Table 6, 561 observations remained after the perpendicular distances were left-truncated 0.2 km and right-truncated 4.0 km, of which 488 were seen by the front observers and 396 by the rear observers; thus 323 were duplicate sightings. The MRDS model assumed point independence because the full independence model showed a lack-of-fit, such that it is assumed that detections made by the front and rear observers are independent except for at distance zero (in effect this is at 0.2 km because of the left truncation). The best detection function was a hazard rate, with no covariates for the DS component, but with perpendicular distance, raw Beaufort Sea state and raw group size (raw because these have not been binned) in the MR component (see Appendix 1 for details on best detection function fit).

Table 6: Sighting seen by primary (front) and secondary (back) observers, and the number of sightings seen by both

Survey	Group size	Primary observer	Secondary observer	Seen by both	Total
2012	1	79	25	77	181
	2	39	8	105	152
	3	2	1	23	26
	4	1	0	4	5
	5	0	0	0	0
	6	0	0	1	1
	7	0	0	0	0
	8	0	0	0	0
2012 total		121	34	210	365
2014	1	46	30	67	143
	2	25	16	57	98
	3	6	2	8	16
	4	1	1	6	8
	5	0	0	3	3
	6	0	1	1	1
	7	0	0	1	1
	8	0	0	1	1
2014 total		78	50	144	272
Total (2012+2014)		199	87	354	637

Figure 4 shows the frequency histograms and fitted detection probability as a function of perpendicular distances for the front and rear observers; Figure 5 for the detections pooled across both front and back observers. Figure 6 shows the conditional detection function plots, which is the probability that one of the observers will see a whale, given that the other observer has seen it. The open circles in Figure 4, Figure 5, and Figure 6 represent the actual sightings – these appear

segregated because of variation in Beaufort sea state and group size (i.e., where this is a higher probability of sighting a whale in calmer seas and with higher group sizes).

For reference, the estimated mean group size was 1.62 with a standard error of 0.03 and the average probability of detection within the surveyed strip was 0.55 with a standard error of 0.024, hence the estimated average effective strip half width (uncorrected for *g(0)*) was 2.4 km. Estimated *g(0)*, which is the average probability of at least one platform detecting a group at "zero" distance from the trackline (a level of perception bias only), was 0.96 (SE = 0.01).

4.2.2 Spatial modelling

Distributions and correlations between the range of spatial and environmental covariates is given in Figure 7. In terms of selecting environmental covariates to include in a habitat-base spatial model, the correlations were small enough to ignore.

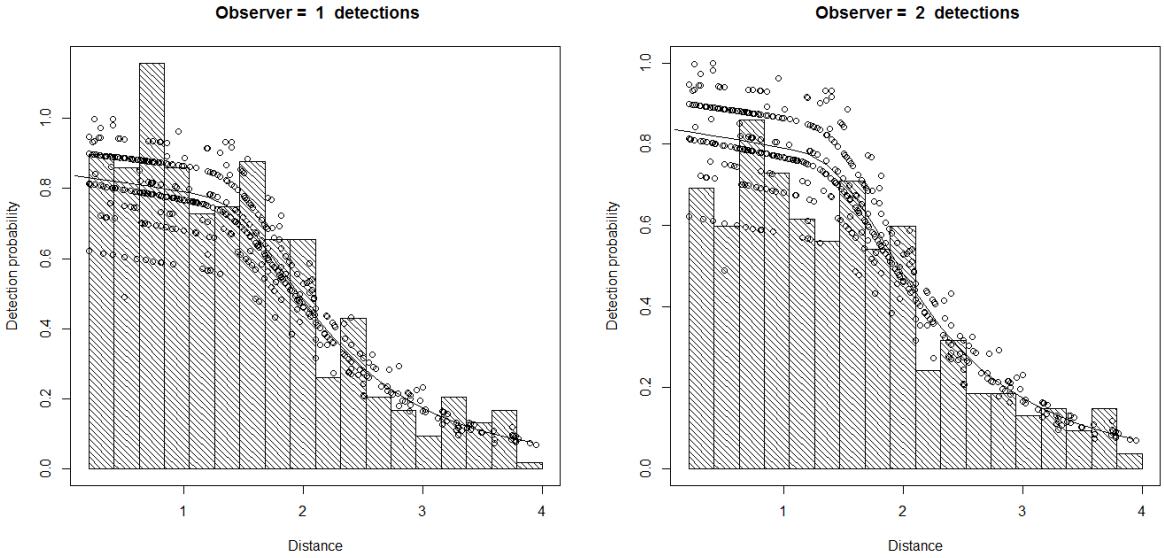


Figure 4: Fitted detection probabilities for the front (primary; left panel) and back (secondary; right panel), with frequency histograms of perpendicular distances to sightings.

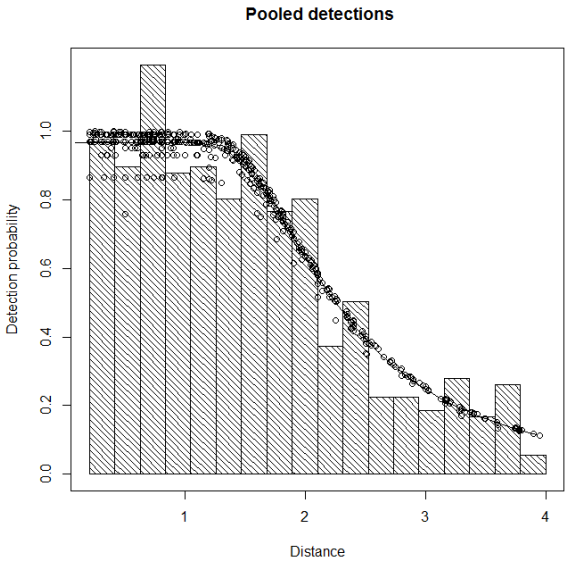


Figure 5: Fitted detection probabilities for pooled sightings (i.e., if either front or back observer saw a sighting), with frequency histograms of perpendicular distances to sightings

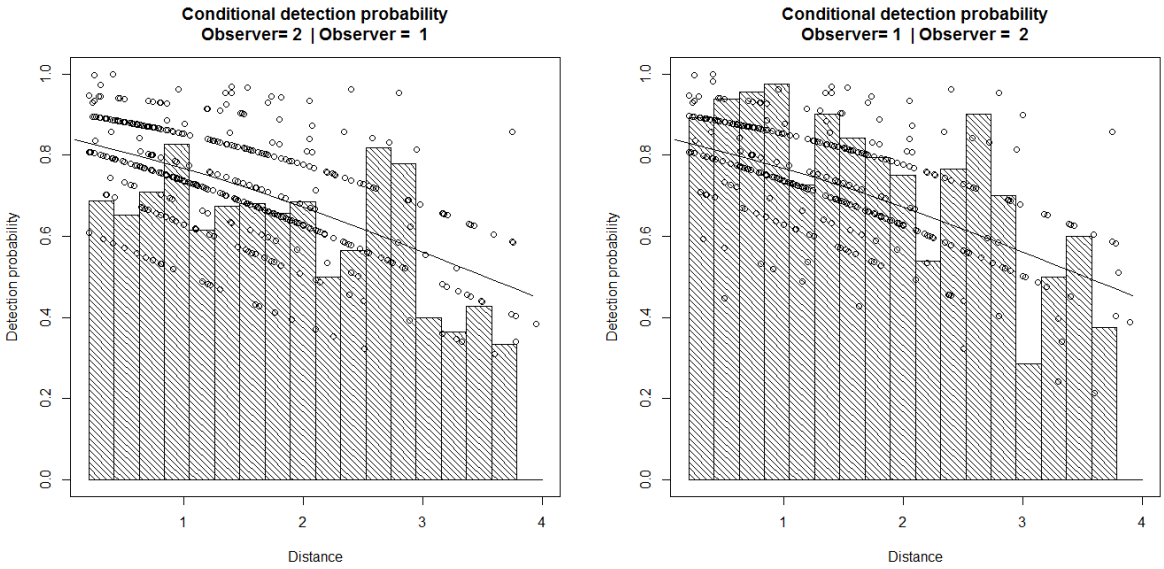


Figure 6: Condition detection plots showing the probability of the (left panel) back observer seeing a sighting given it was seen by the front observer, and (right panel) front observer seeing a sighting given it was seen by the back observer

Spatial models, based on a projected 'easting' and 'northing' value (transformed from longitude and latitude using the Albers equal area projection), for animal-wise densities derived from all sightings, for 2012 and 2014 are given in Figure 8. Spatial models predicting the distribution of densities of animals in groups accompanying calves, for both survey years, respectively, are in Figure 9. The distribution of sightings, and estimated group size, are provided as a form of model validation. Separate spatial models were fitted for each survey year, and for all animals, or for animals in groups with calves. Details of the tensor-product smoothes in these plots are given in Appendix 2.

The best GAM, or density surface model, describing the distribution of humpback whale densities (all animals) with various physiographic and environmental covariates is displayed in Figure 10. This model is summarised in Appendix 3. This GAM based on a tensor product smooth between bathymetry and sea surface temperature, and individual thin-plate smoothes of sea surface temperature and sea surface height anomaly. The addition of a 'survey' factor did not significantly improve the GAM fit, so it was not retained in the model. Owing to only a small amount of survey effort in bathymetric values of 90 m and deeper (only 122 km of a total of 6650 km across both survey years), no density predictions were made for waters deeper than 90 m. The surface density model based on bathymetry, sea surface temperature and height anomalies tallies well across space, as predicted by the space only GAMs described above. The predicted distribution of densities of animals accompanying calves is given in Figure 11; this GAM is summarised in Appendix 3.

Additional analysis of other covariates is presented in Appendix 4.

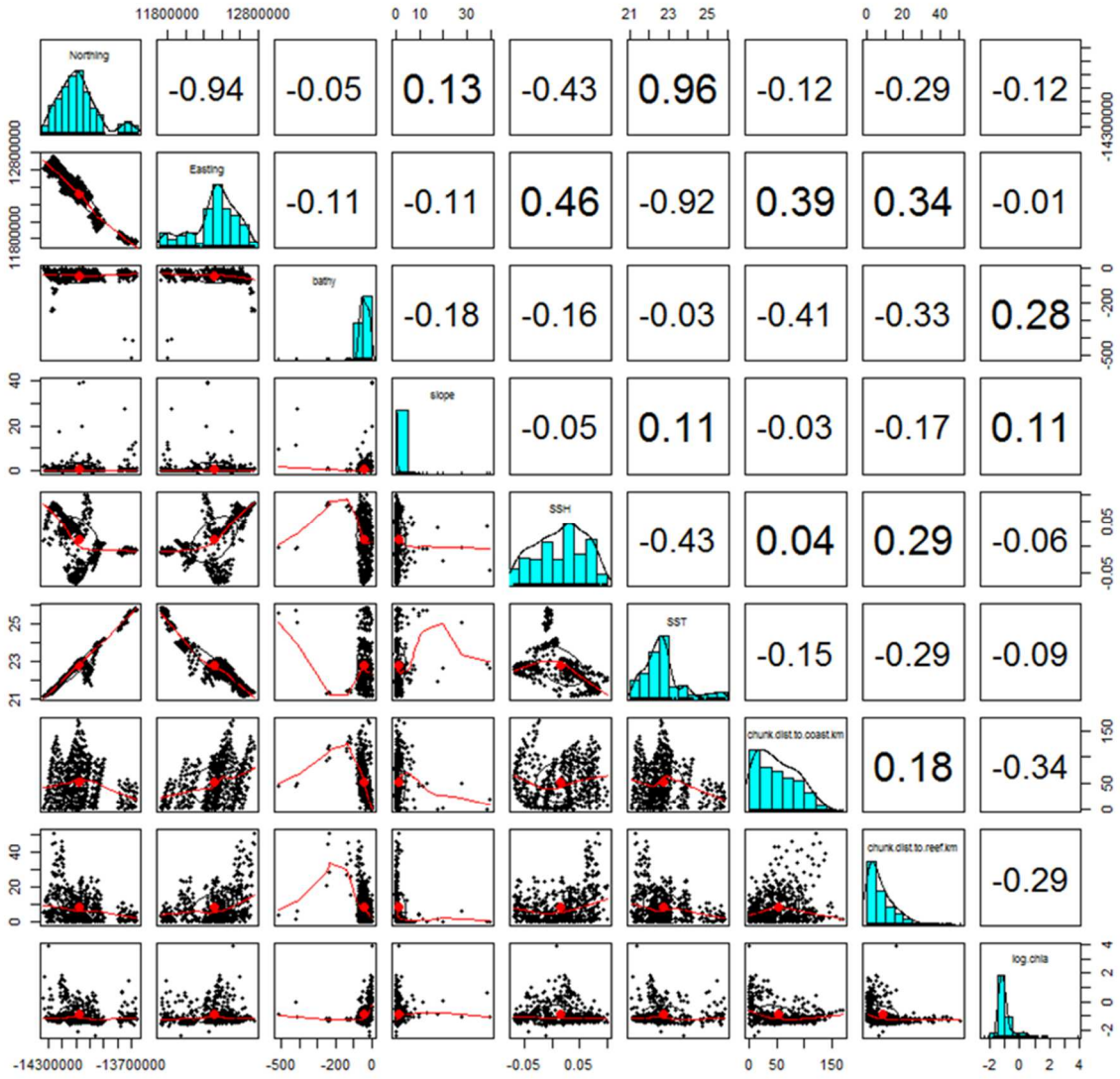


Figure 7: Matrix of scatterplots of spatial and environmental covariates, as inferred at the middle of each 10 km along-track segment. The spatial covariates are represented by an 'Easting' and 'Northing' value, which arose from an Albers equal area projection to the latitude and longitude data. Pearson correlation coefficients given in the upper-right of the matrix

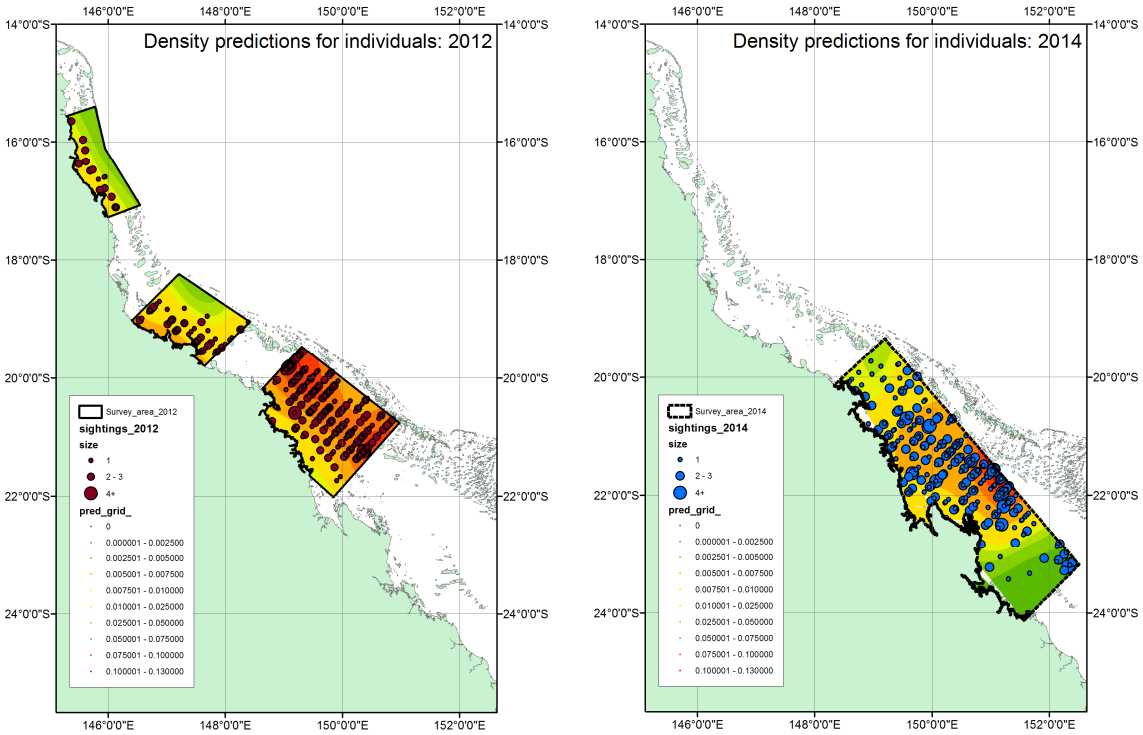


Figure 8: Predicted densities of humpback whales (all animals) within the surveyed areas for the 2012 and 2014 surveys, based on a 'spatial' density surface model. Animal density scale given in the legend. Distribution of sightings given as a form of model validation

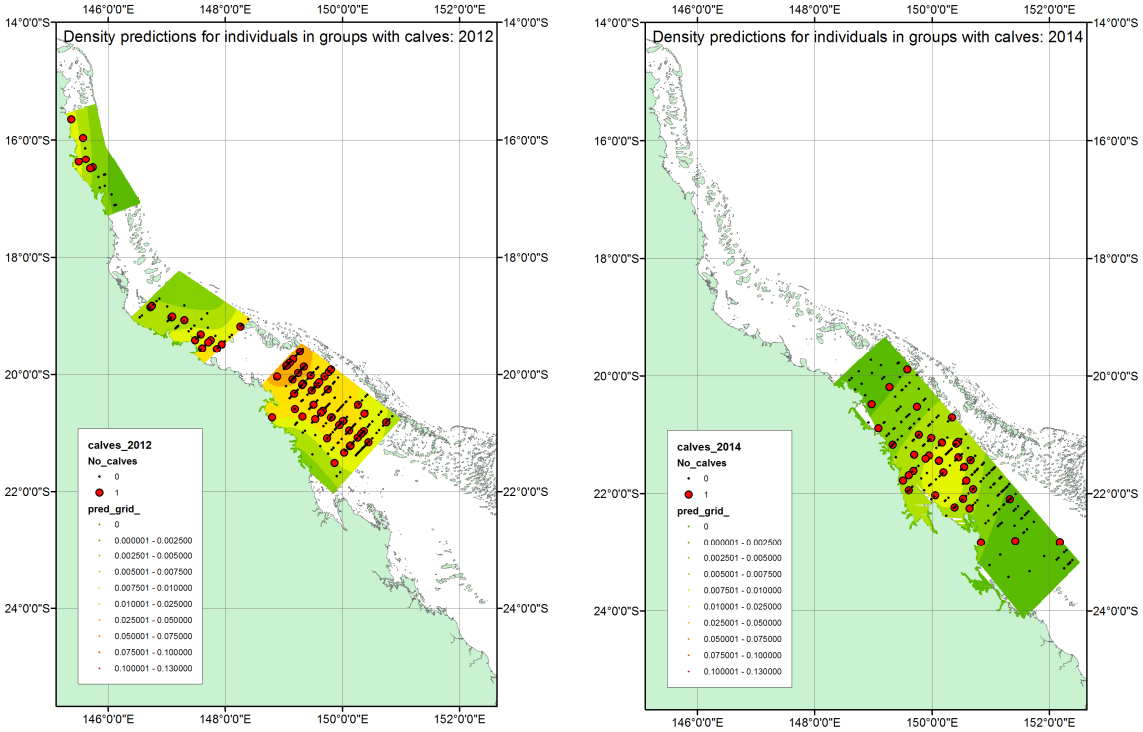


Figure 9: Predicted densities of humpback whales in groups accompanying calves within the surveyed areas for the 2012 and 2014 surveys, based on a 'spatial' density surface model. Animal density scale given in the legend. Distribution of sightings given as a form of model validation

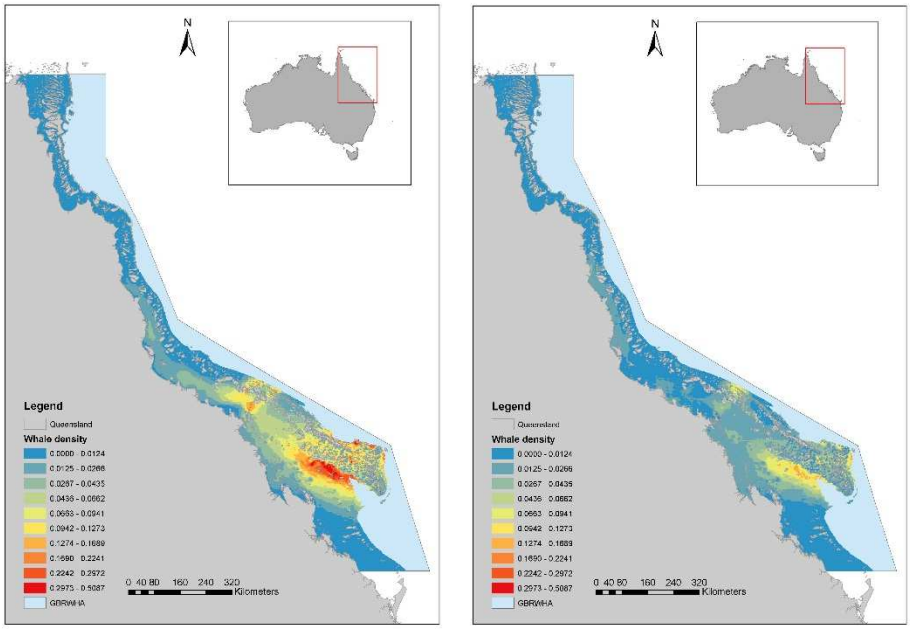


Figure 10: Distribution of densities of humpback whales (all animals), extrapolated throughout the GBR, using a density surface model based on selecting potentially influential physiographic and environmental covariates, as they were recorded in August 2012 and September 2014

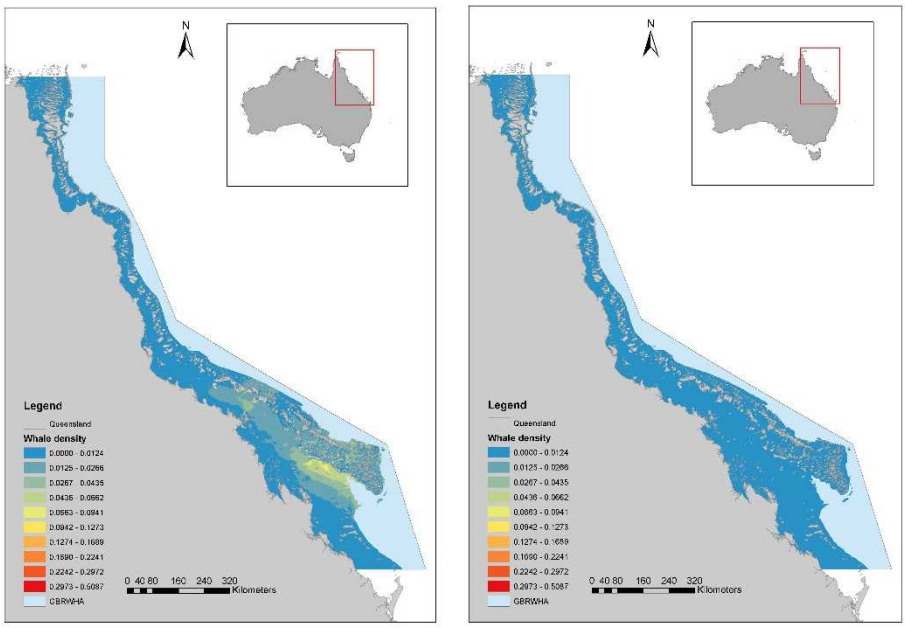


Figure 11: Distribution of densities of humpback whales in groups accompanying calves, extrapolated throughout the GBR, using a density surface model based on selecting potentially influential physiographic and environmental covariates, as they were recorded in August 2012 and September 2014

4.3 Shipping data

4.3.1 AMSA AIS data

The AMSA CTS data of vessel movements was successfully imported and processed. The total counts of data sizes at each stage of the processing/analysis are given in Table 7. To estimate future computational requirements we summarise storage (Table 8). The total storage for the data and processing was around 12GB and to process a typical year start to finish took about a day.

Table 7: Summary of data quantities at each stage of the processing

Year	Mth	Raw data	Filtered data points	Transects	
		Number	Number	Number	Dist. (km)
2012	Jul	390,274	245,226		
	Aug	392,737	229,077		
	Sep	400,626	203,991		
	Total	1,183,637	678,294	655,324	1,342,556
2013	Jul	377,435	248,591		
	Aug	362,147	213,004		
	Sep	565,719	268,624		
	Total	1,305,301	730,219	715,409	1,358,744
2014	Jul	456,035	264,179		
	Aug	498,120	313,910		
	Sep	494,108	313,910		
	Total	1,448,263	891,999	844,589	1,602,881
	TOTAL	**3,939,201**	**2,301,512**	**2,215,322**	**4,304,181**

Looking at basic summary statistics of the raw data, we saw a slight increase in amount of shipping data in later years[9] (Figure 12) and no discernible difference within season. However, there was a large peak of data in September 2013 the reason for which is not clear. It should be noted however that these changes could be due to refinement in the AIS system reporting (e.g., more frequent polling, more reporting vessels) than a reflection of actual vessel traffic changes.

The majority of class A vessels (>= 80 m in length) were cargo vessels, followed distantly by Tankers and a small number of passenger vessels (Figure 13).

In terms of the size of the vessels, the length ranged from 80 m (our minimum cut-off) and 300 m (Figure 14) with a mean of 205m and a median of 222m. Vessel beam ranged between 10 and 50m (Figure 15) with a mean and median of 32m. These histograms/statistics are weighted by time in the GBR, hence there will be a bias toward vessels with lower speeds. These plots can be redone, if required, weighted by distance traversed to remove this bias.

[9] This calculation was based on a raw count of data assuming all vessels were polled at 5min intervals so is an approximation only.

The modes in the data may represent a combination of a certain size vessel making a number of voyages within the season and also that there are certain types of sized ships based on international canal requirements. The draught information (Figure 16) was a bit patchy and seemed to include unrealistic draughts. If in future it is decided to use vessel draught in the risk calculations, some further data cleaning will be required.

Vessel speed was nicely distributed (Figure 17) with a median at around 12 knots. Again this is just a preliminary data exploration and will be biased toward slower vessels.

Table 8: Summary of data file sizes at each stage of the processing

Year	Month	Raw (MB)	Filtered (MB)	Transects (MB)
2012	Jul	931	123	
	Aug	977	125	
	Sep	851	127	
	Total	2759	375	255
2013	Jul	1000	128	
	Aug	1010	112	
	Sep	1297	196	
	Total	3307	436	278
2014	Jul	1115	141	
	Aug	1171	155	
	Sep	1137	154	
	Total	3423	450	343
	TOTAL	9489	1261	876

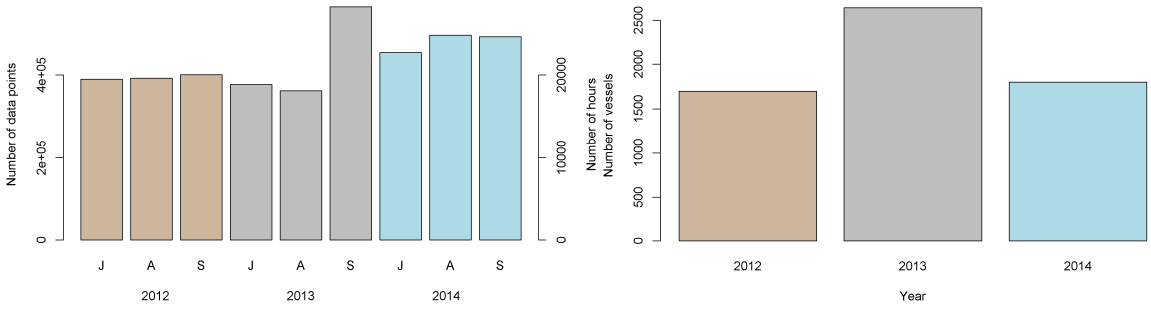


Figure 12: Totals per month counts (left axis) of cargo, tanker and passenger vessels of length >= 80 metres in the GBR and total vessel hours (right axis).

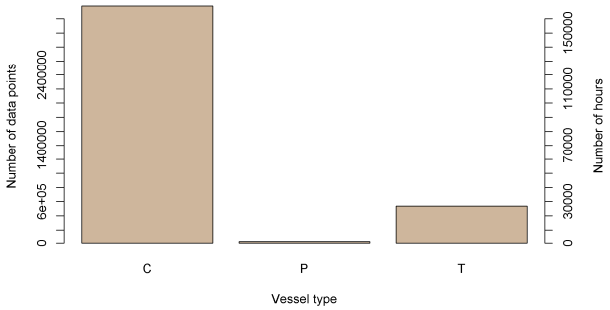


Figure 13: Total count data points (left axis) of cargo, tanker and passenger vessels of length >= 80metres in the GBR for winter 2012-2014 and total vessel hours (right axis).

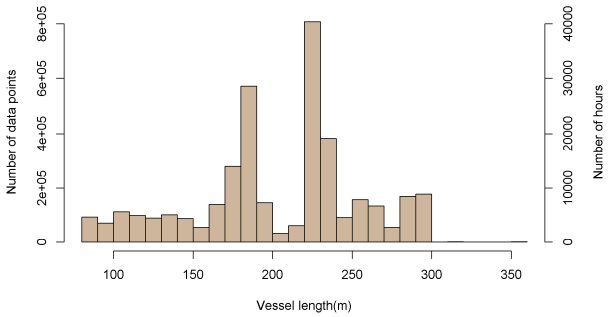


Figure 14: Distribution of vessel lengths of cargo, tanker and passenger vessels of length >= 80metres in the GBR in winter in terms of number of data points (left axis) and total vessel hours (right axis).

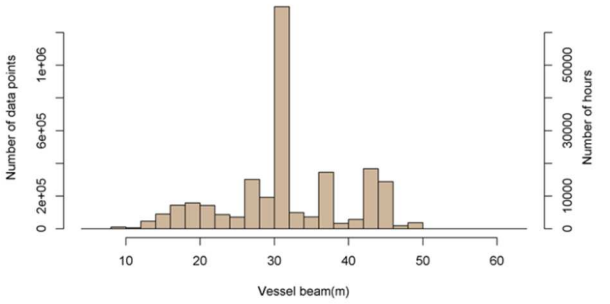


Figure 15: Distribution of vessel beam of cargo, tanker and passenger vessels of length >= 80metres in the GBR in winter in terms of number of data points (left axis) and total vessel hours (right axis).

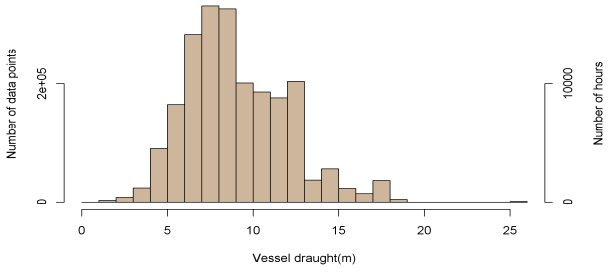


Figure 16: Distribution of vessel draught of cargo, tanker and passenger vessels of length >= 80metres in the GBR in winter in terms of number of data points (left axis) and total vessel hours (right axis).

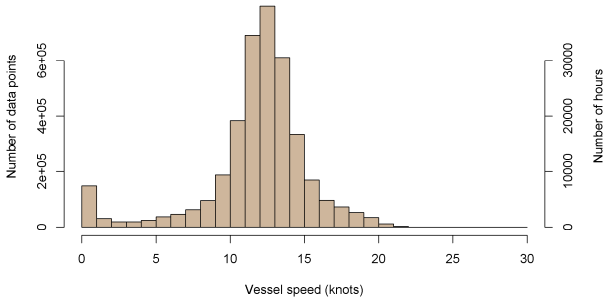


Figure 17: Distribution of vessel speed of cargo, tanker and passenger vessels of length >= 80metres in the GBR in winter in terms of number of data points (left axis) and total vessel hours (right axis). Note: This will be biased toward lower speeds.

4.3.2 Shipping traffic density

Once the data was filtered and processed, the result is a summary of vessel activity at 1 km^2 grid cells across the GBR. The main measure used in ship strike analysis is total distance traversed by vessels within the grid cell (Figure 19, Figure 20 and Figure 21).

4.3.3 Other information/outputs

4.3.3.1 Vessel Type

We can also split the results by vessel type (i.e., cargo, passenger or tanker) and look at spatial differences in distribution (Figure 18) and therefore potentially assess relative risk of each vessel type.

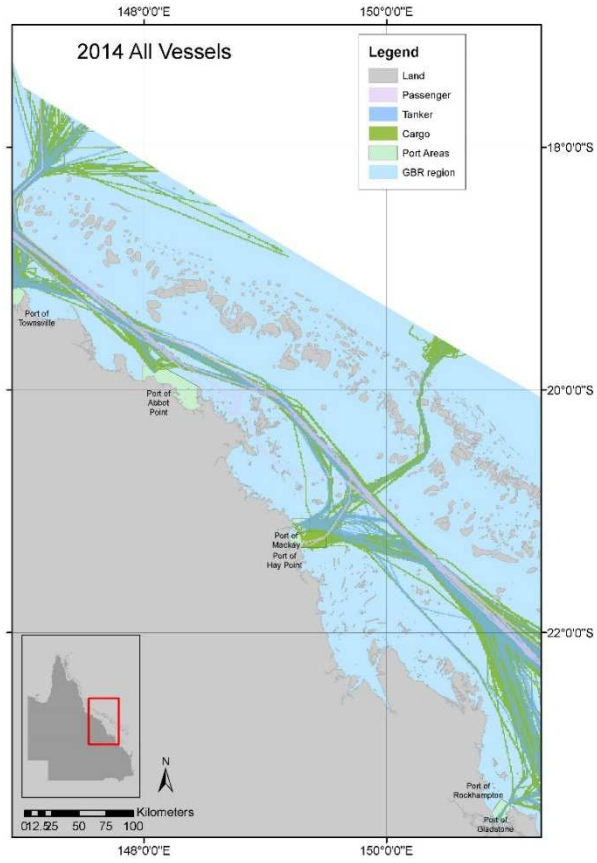


Figure 18: Example of the spatial separation between cargo, passenger and tanker vessels

4.3.3.2 Vessels effected

One useful statistic we summarise for each 1 km^2 grid cell is the number of unique vessels that have passed through each grid cell (Figure 22). This can be informative of the impact of any management restrictions. We can count vessels based on MMSI (as per Figure 22) or it may make more sense to count based on IMO number or transits.

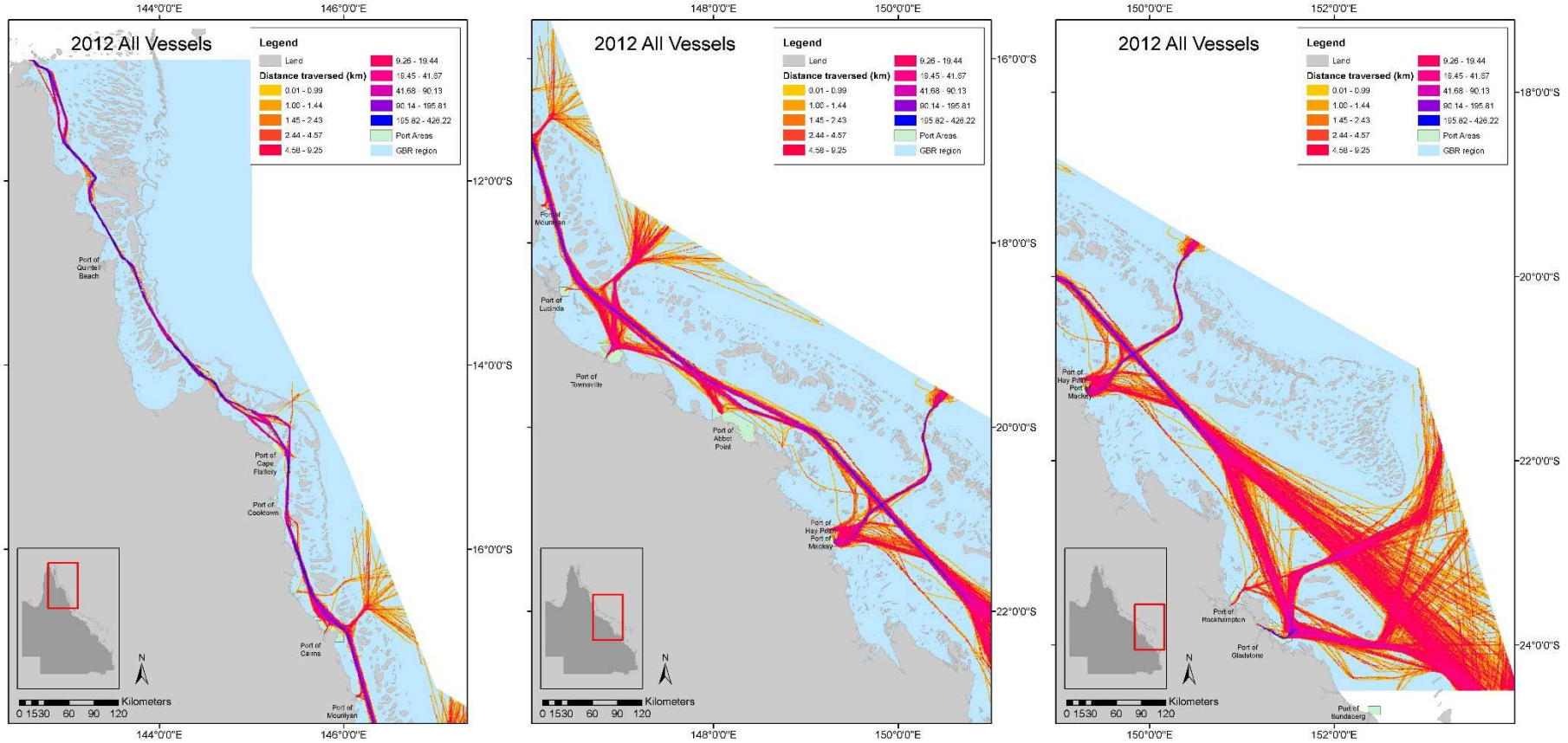


Figure 19: Winter 2012 distance traversed in the Northern, Central and Southern extent of the GBR

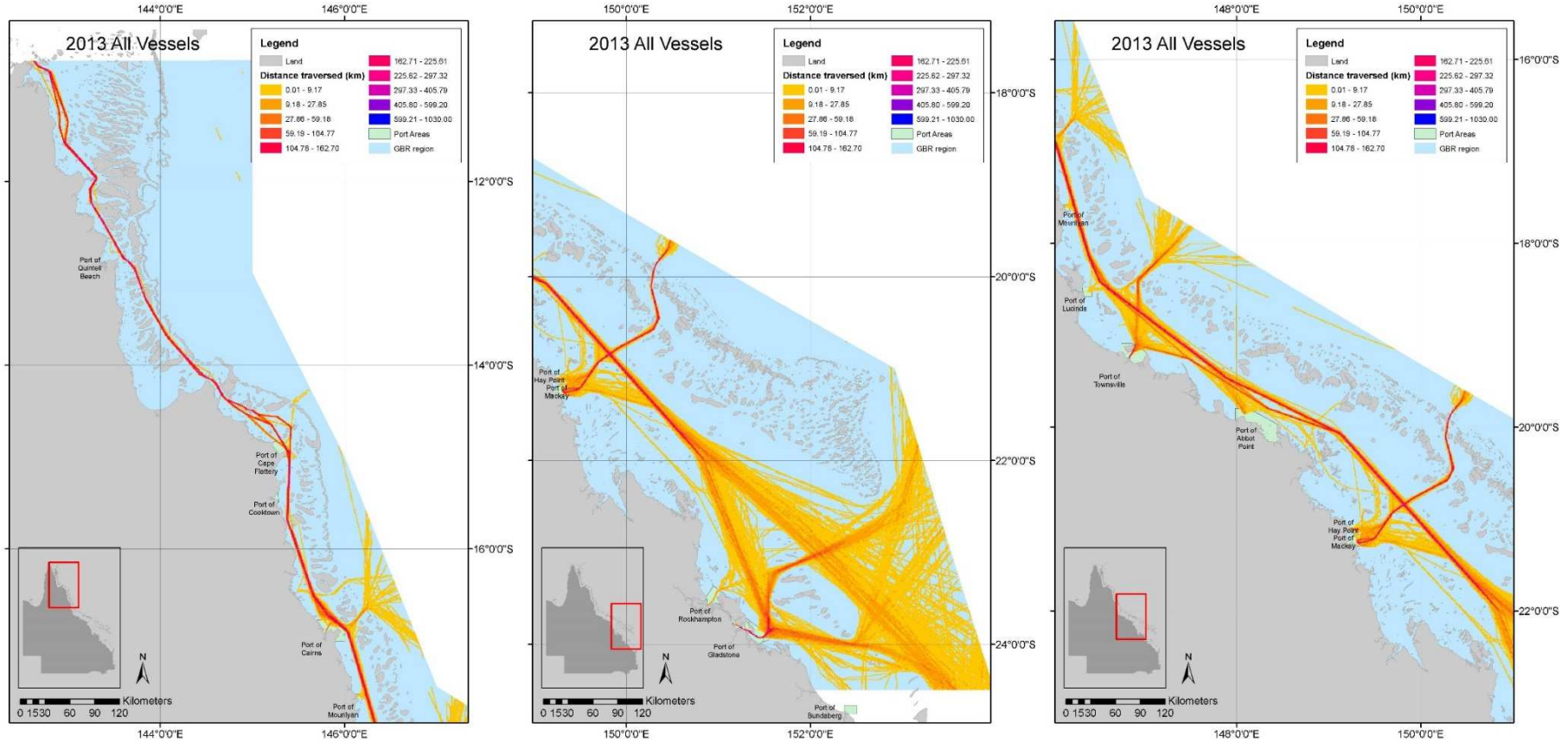


Figure 20: Winter 2013 distance traversed in the Northern, Central and Southern extent of the GBR

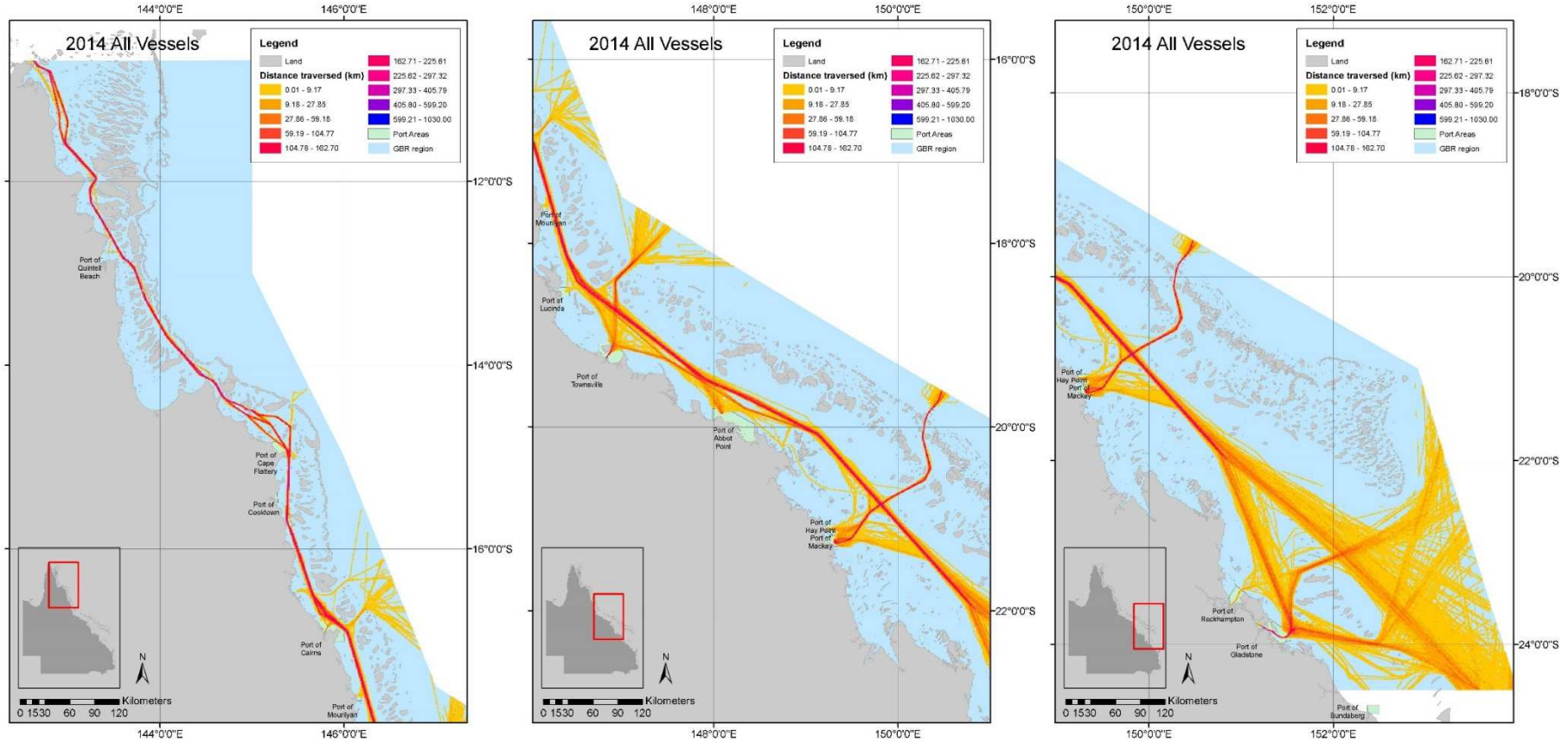


Figure 21: Winter 2014 distance traversed in the Northern, Central and Southern extent of the GBR

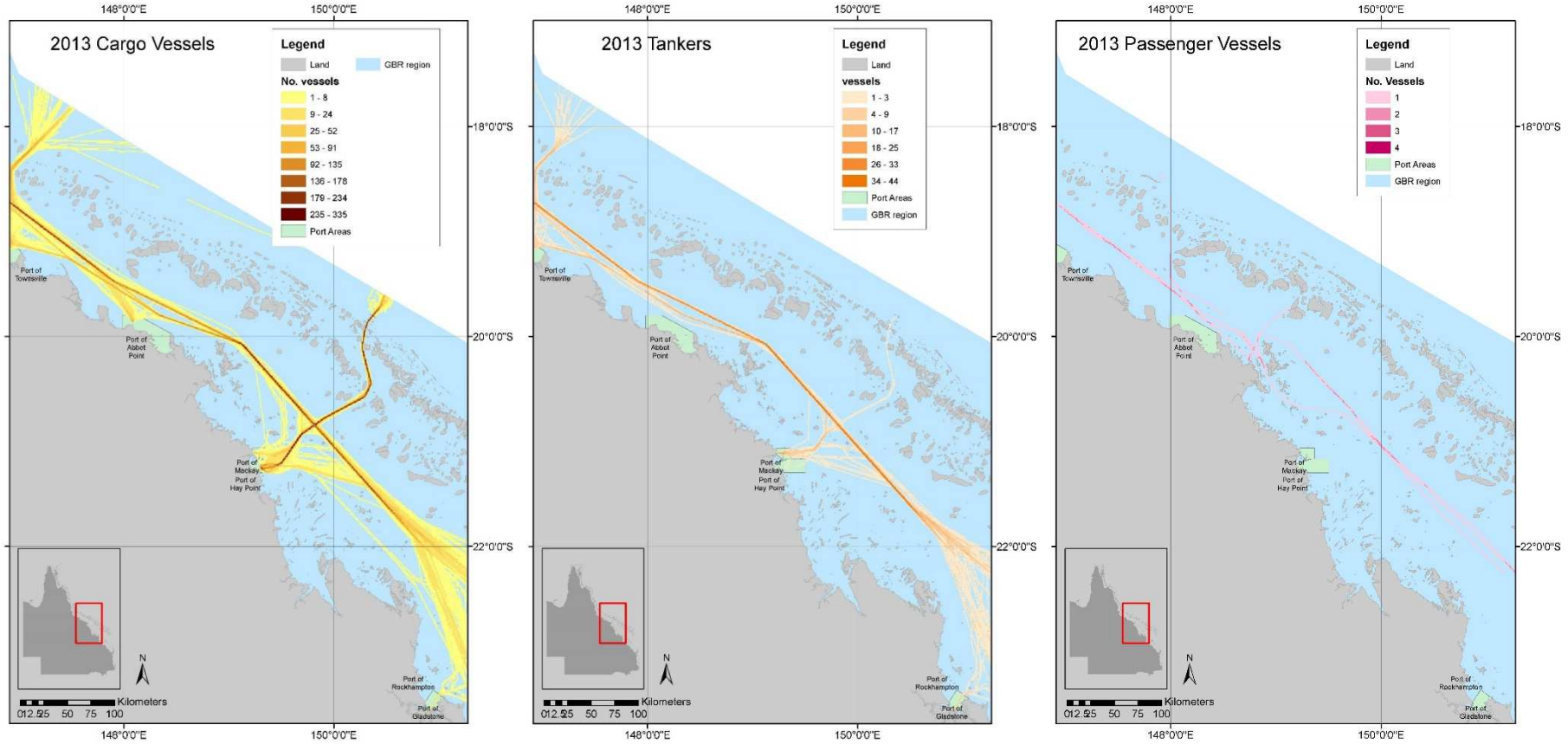


Figure 22: Example of distinguishing the number of unique vessels (based on MMSI) that passed thru each grid cell

4.4 Other data

While we didn't specifically set out to explicitly investigate environmental correlates with whale distribution, as part of the spatial modelling exercise, we found that the most significant parameters in describing humpback whale distribution and density were depth and SST. These were by far and away the biggest contributors to describing the distribution of humpbacks within the GBRWHA. There is further details of these analysis provided in Appendix 6.

4.5 Statistical analysis

4.5.1 Relative risk maps

The co-occurrence risk was calculated as per section 3.4.2.1. So for each combination of whale year (winter 2012 and 2014) and vessel data (winter 2012, 2013 and 2014) we matched the whale and shipping density data for each 1 km^2 grid cells (Figure 25, Figure 26) and then estimated co-occurrence. This approach does give some indication of temporal uncertainty which we report as the range of the resulting 6 replicate grid cell values (arising from the 3x2 combinations) divided by their mean (herein called 'standardised range'). In addition, as a proof of concept we provided the index of expected number of fatalities (section 3.4.1.2).

4.5.2 Co-occurrence Comparisons

We compared the total sum of co-occurrence for various subsets of the data[10] with a summary of results shown in Figure 23 and Figure 24. As would be expected given their higher number of vessel movements, cargo vessels represent by far the largest proportion of relative risk. This result does not provide evidence to say individual single cargo vessels pose more of a risk as if we look at the average risk per vessel km travelled in each type, then individual cargo and tanker vessels give similar results per km and passenger vessels show a slightly higher per vessel km risk. However, with only 15 passenger vessels (> 80 m) in our analysis this result may be a result of the small sample size. If the difference in relative risk per individual vessel km is real then since we do not include vessel speed or size in the co-occurrence index, this difference would be driven by simply more spatial co-occurrence.

There were inter-annual differences in the number of vessel movements and also humpback density which meant that the relative risk for 2012 was considerably higher than for 2014.

While we did not include vessel speed in this component of the modelling, it is possible that if different vessel types have different speed characteristics, therefore this simplistic model wouldn't be accurately capturing the true relative risk. The probabilistic model outlined in section 3.4.1.2 would allow the inclusion of vessel speed. Preliminary results using the relative index of the expected number of fatalities, that incorporate vessel speed show a slightly increased per km risk for tankers compared to cargo vessels and a larger increase again for passenger vessels. However, it should be noted that even if after further investigation it does prove to be true there is a vessel type difference per km, the total cumulative exposure is what is important, and in this case cargo vessel contribute predominantly to the overall risk due to the number of vessels.

With respect to the relative risk for groups with and without calves, the modelling indicates that the non-calf groups have a higher relative overall cumulative risk. However, this is primarily driven by there being significantly more non-calf groups than calf-groups. If we look at the standardised risk per whale in each group, then the risk is similar and possibly slightly higher for groups with calves. Again, since our model doesn't consider the difference in the risk of being struck between the two

[10] The co-occurrence metric is summed to calculate a total for an area. This can be shown to be correct but due to time/space we shall not include in the report.

groups[11] then if this slight apparent difference is real, it would correspond to slightly higher proportion of the calf groups having co-occurrence with vessels. Preliminary results using the relative index of the expected number of fatalities show no differences between whale groups

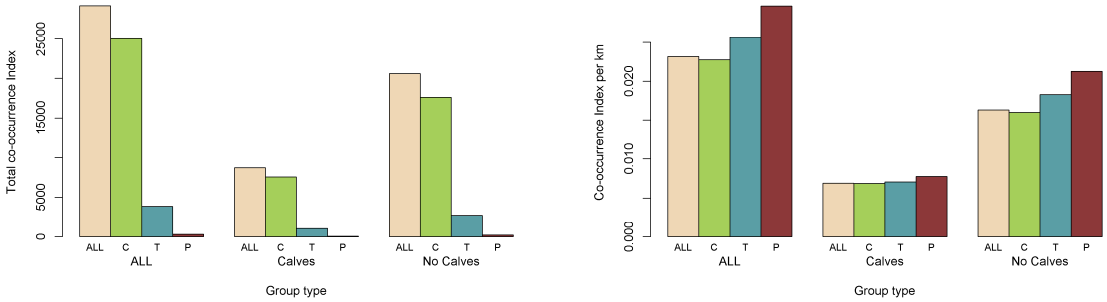


Figure 23: Comparison of total co-occurrence risk for each vessel type (left) and standardised by total number of vessels km travelled to give average risk per vessel km for each vessel type (right). (Note: within class A >=80 m length)

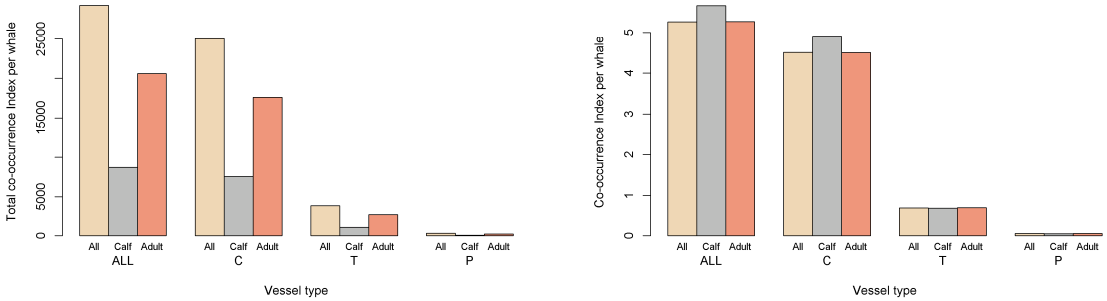


Figure 24: Comparison of total co-occurrence for different whale group types (left) and standardised by total number of whales in each group type to give average risk per whale for each whale group type (right).

4.5.2.1 Broad scale Comparisons

The results of the large-scale comparisons are shown in Figure 27 and Figure 28 with the GBRWHA divided into 150x150km grid cells. The model found that the central band between Rockhampton and Mackay to have the largest relative risk. However, it is important to note that the Southern area outside Gladstone contains deep water, for which it was not possible to reliably predict whale density and so the results for those grid cells will underestimate risk.

Figure 27 shows the plot but broken down by vessel type. The same general wide scale pattern is present. Also included are broad scale plots based on 50 x 50km^2 grid cells to give further detail but still remain at the broad scale (See Figure 29).

[11] for example, there potentially could be differences in dive behaviour and vessel avoidance between calf and non-calf groups

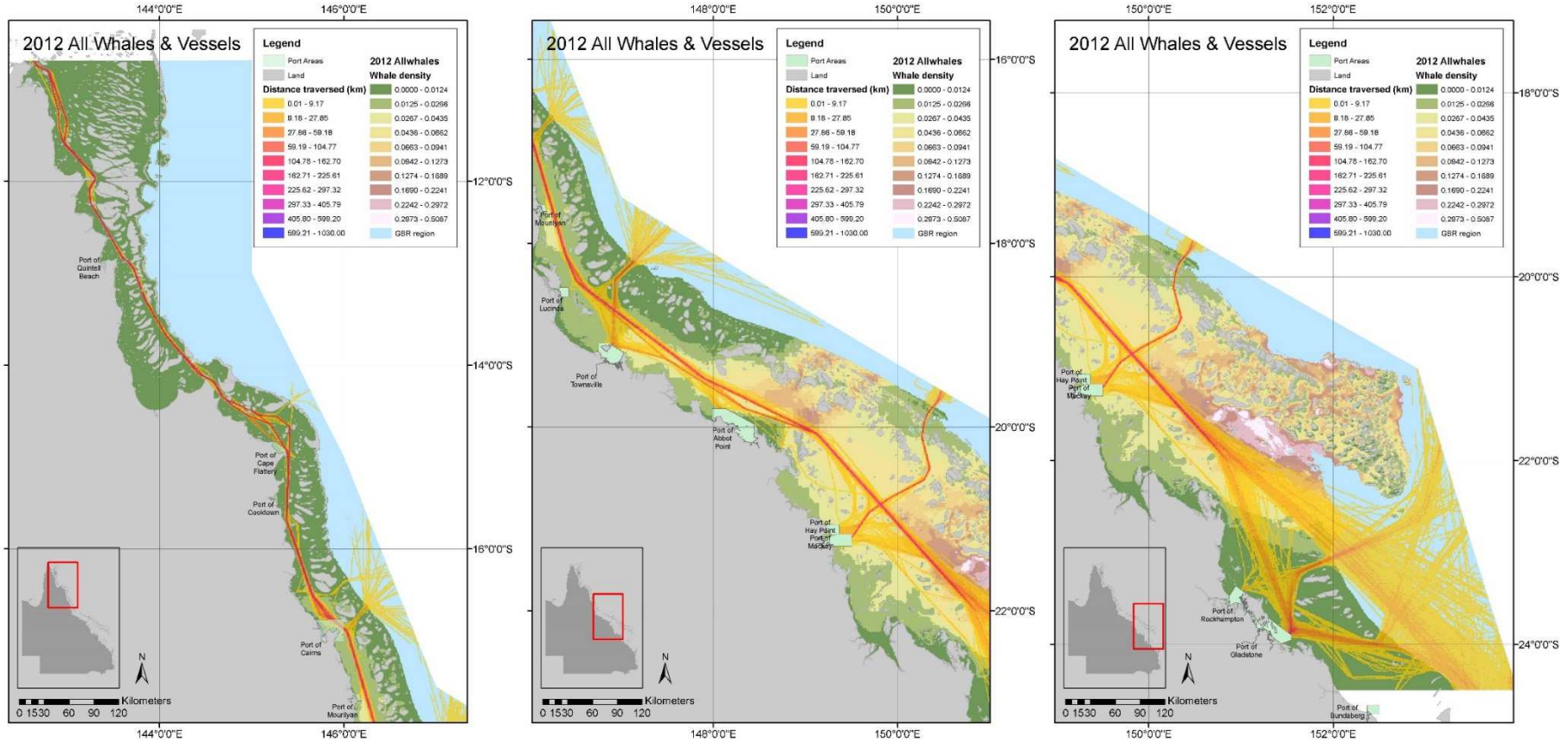


Figure 25: Simple overlay of 2012 whale data and all shipping distance traversed data

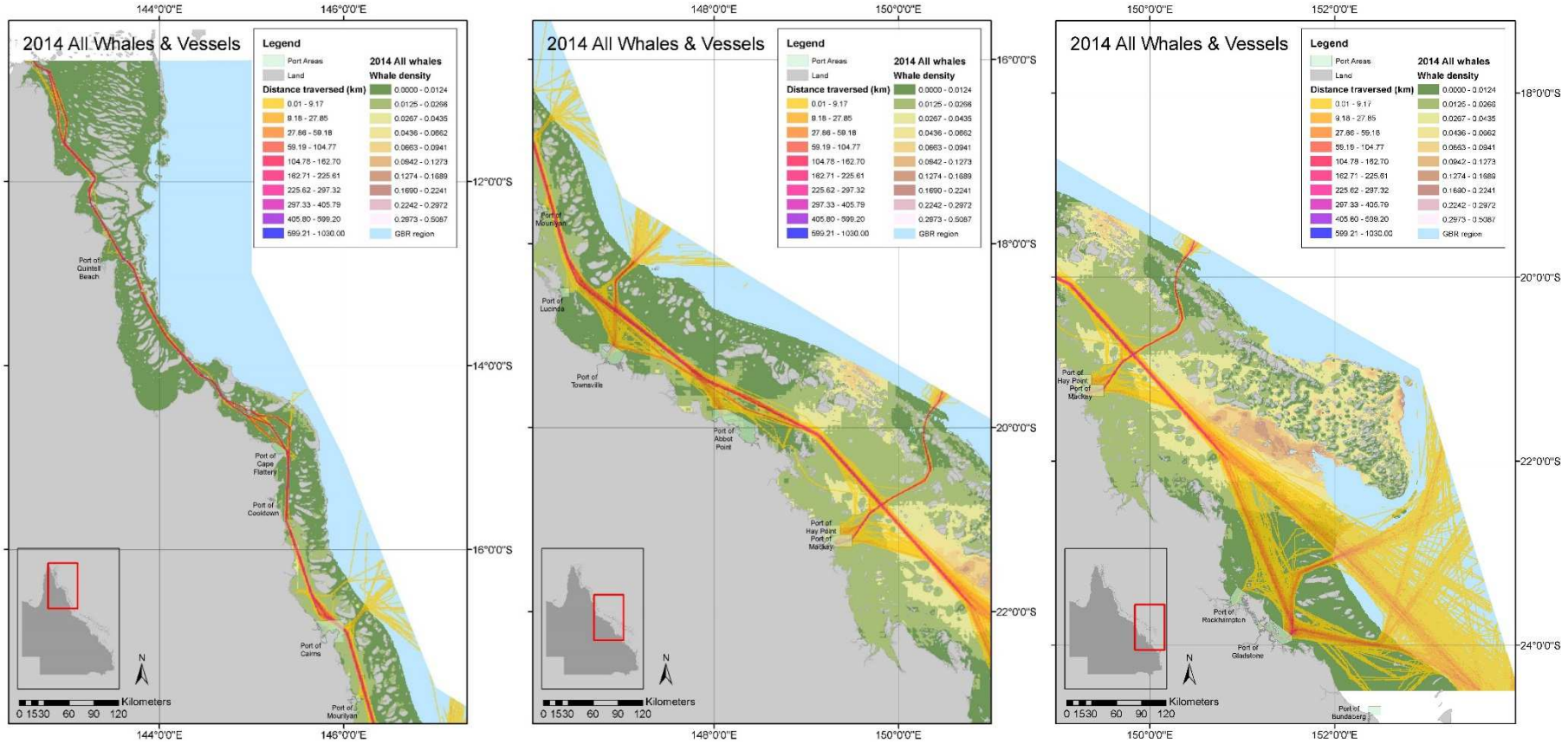


Figure 26: Simple overlay of 2014 whale data and all shipping distance traversed data

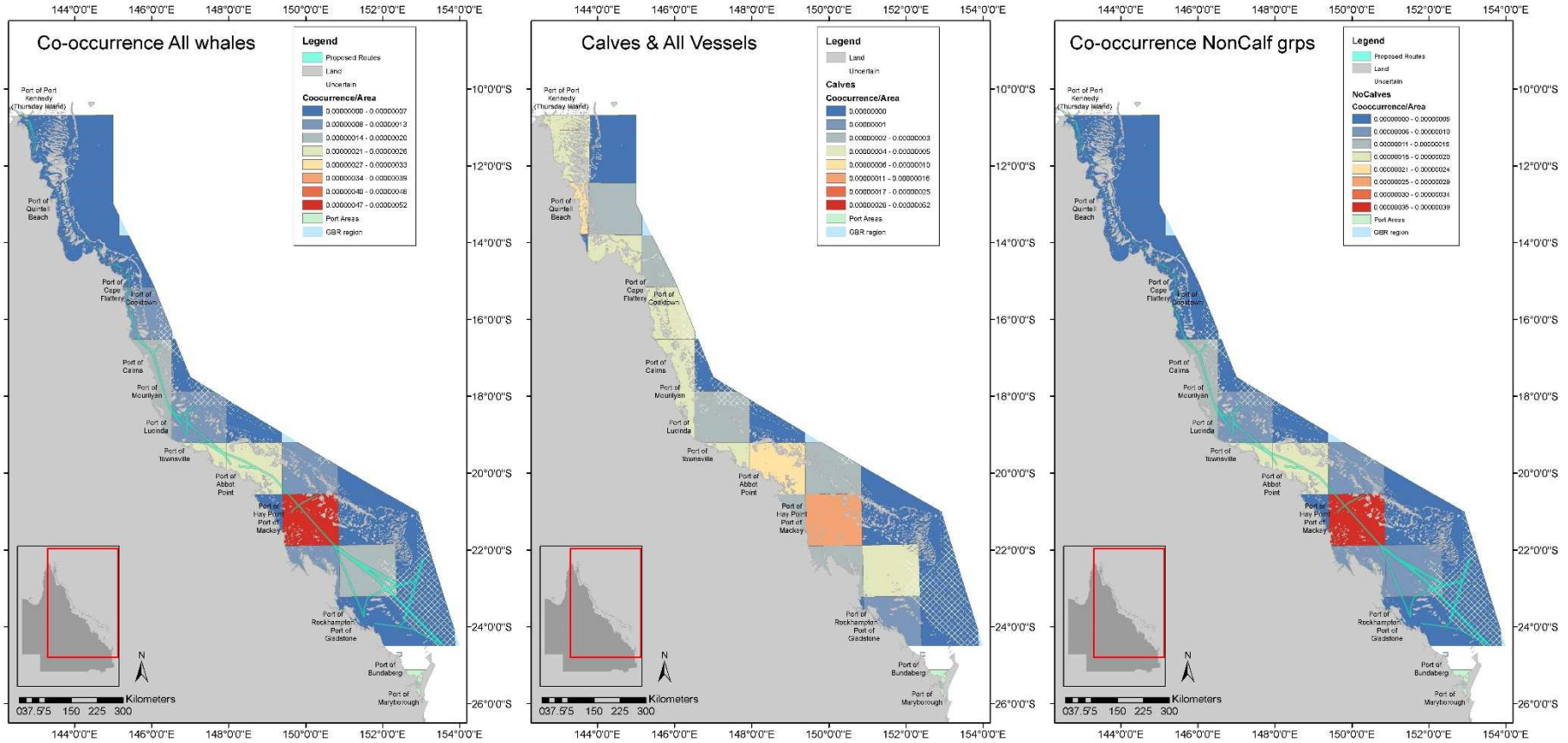


Figure 27: Large-scale plot (150x150km grid cells) of the co-occurrence index (standardised by area) for various whale group types with all vessels. Note: The uncertain areas where no estimation is done denoted by the hatching.

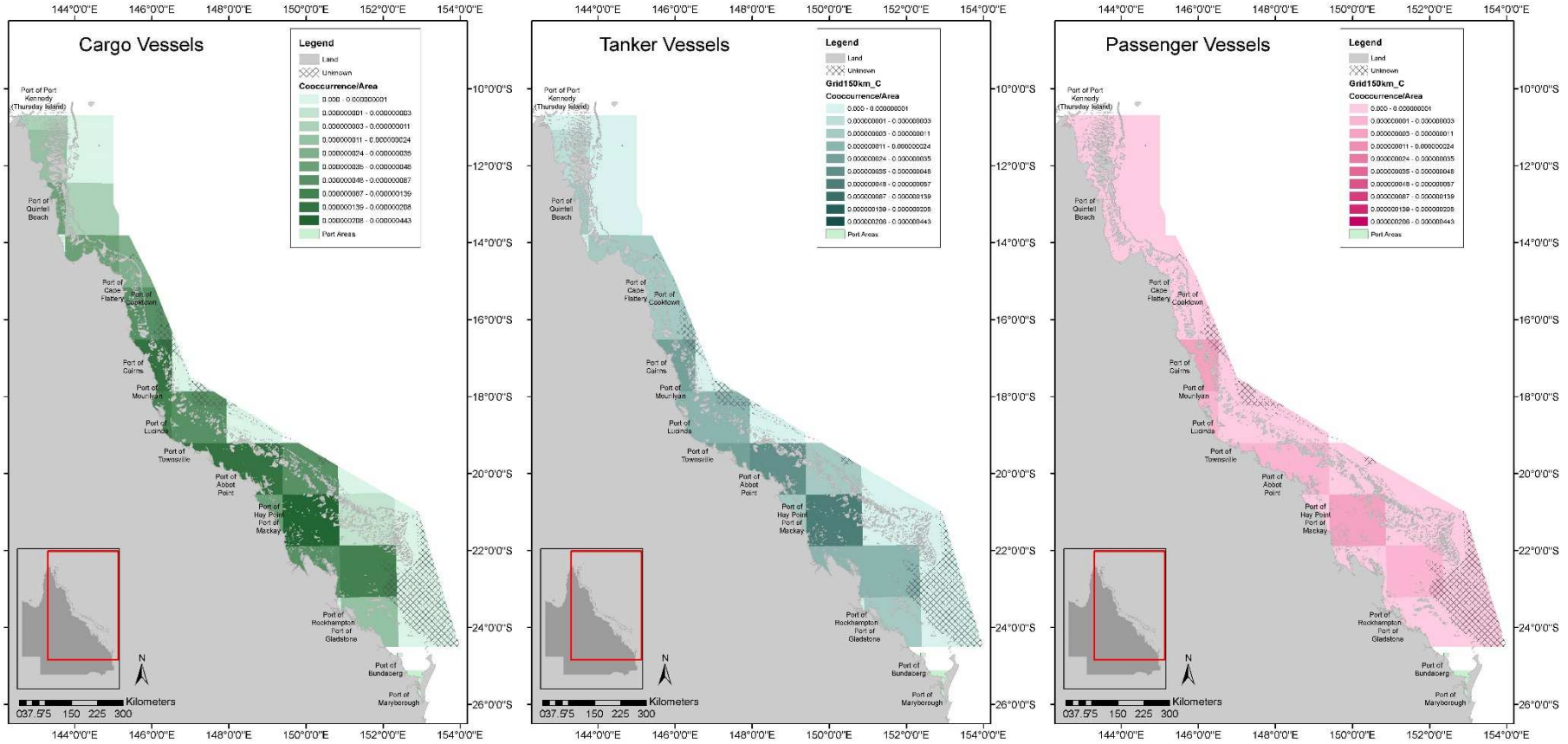


Figure 28: Large-scale plot (150x150km grid cells) of the co-occurrence index (standardised by area) for various vessel types with all whales. Note: The uncertain areas where no estimation is done denoted by the hatching.

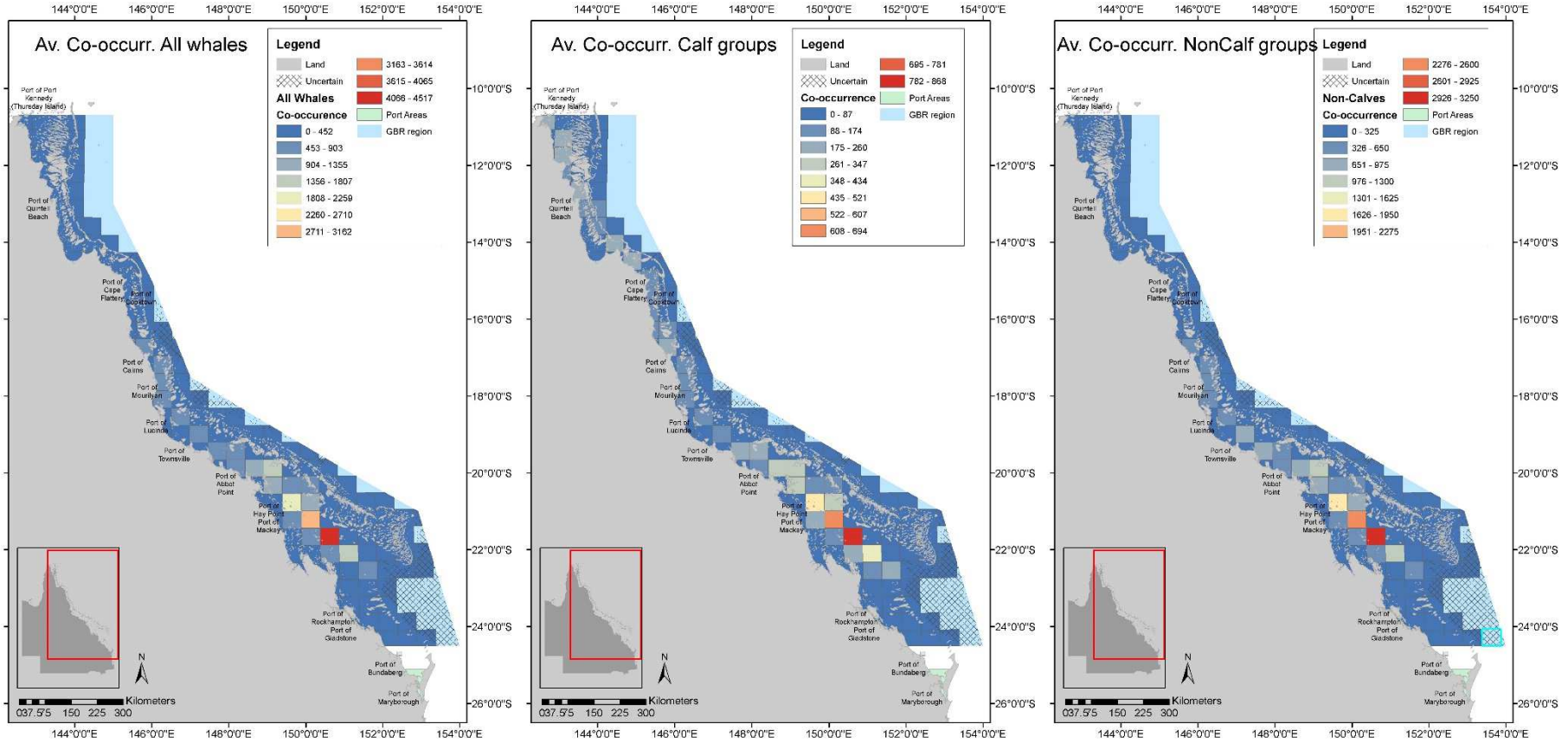


Figure 29: Medium-scale 50 x 50km grid cell plot of the co-occurrence index for various vessel types and all whales. Note: The uncertainty (hatched area) so the total in those grid cells corresponds to the shallower unhatched area only

4.5.2.2 Fine-scale Co-occurrence risk maps

Finally we plotted the various data and its subsets on a 1x1 km grid for the GBR, for example as shown in Figure 31: and Figure 34:. These figures show the co-occurrence index for all whales and calf groups against all vessels, for various extents in the GBR. We also include a map of the standardised range to give an indication of the variation across all year combinations that was seen.

4.5.2.3 Other Co-occurrence measures

As discussed in Section 3.4.1.1 there are more than one way to characterise co-occurrence, in particular the data can be transformed to alleviate some of the issues with the disparate scales of shipping intensity and whale density. Some preliminary results of two approaches are given in Appendix 6. In the first option, we classified the shipping data into seven categories and the whale data into eight categories using the standard deviation (the difference in the number of categories for each data set was driven by their distributions. We then multiplied the categories, resulting in a map that shows the overlap throughout areas with shipping and whale data. In the second option (high co-occurence), we have simply highlighted where the highest categories of shipping traffic overlap with the highest categories of whale densities (5 or above on both whales and ships).

4.5.3 Index of expected fatality maps

As a proof of concept we developed the relative index of the expected number of fatalities (section 3.4.1.2) and produced maps (Figure 35:). While this is useful, further development is required to refine some of the approximations/assumptions.

Overall this approach provides similar results to the broad scale maps using the co-occurrence index notably there are some fine scale differences. The advantage of this approach is that this measure considers vessel speed so is a better metric to compare relative risk from different vessel types and can easily accommodate difference in strike and fatality heterogeneity (e.g., differences between vessel types and/or whale group types).

4.5.4 Projected future risk increase

Based on the calculations in section 3.4.1.2, we calculated the increase in risk based on a 10.5% annual increase in whale numbers (Noad et al. 2011), assuming the South-East Queensland migratory population growth translates to a similar growth in the GBR. This was done for various shipping traffic growth rates and projected forward to 2024, see Figure 30.

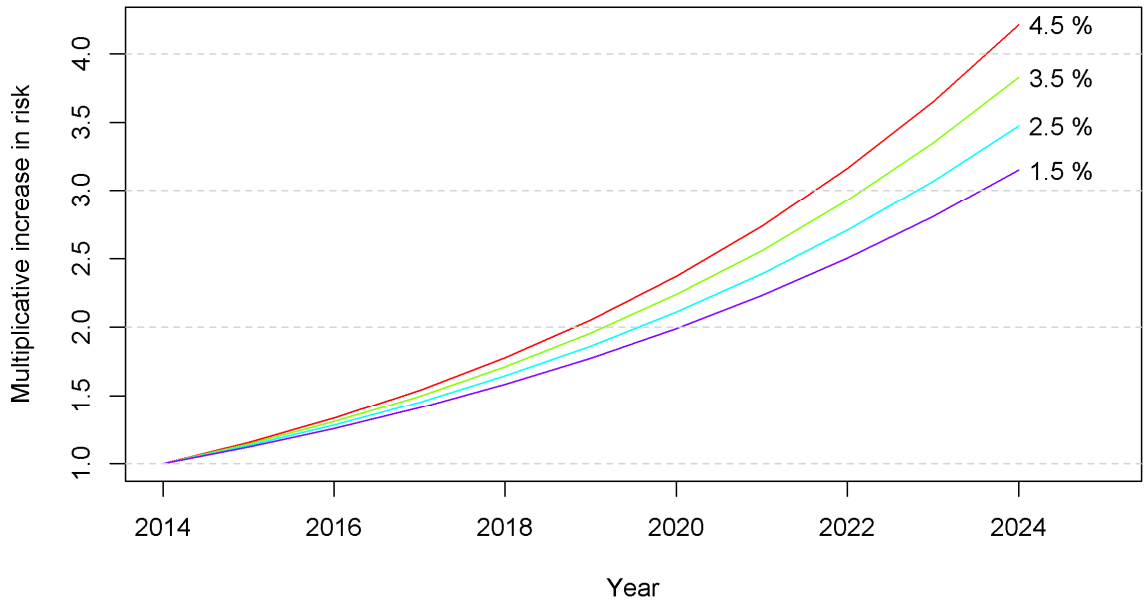


Figure 30: Projected increase in relative co-occurrence risk based on an annual 10.5% whale increase and various projected annual shipping traffic increases (coloured lines)

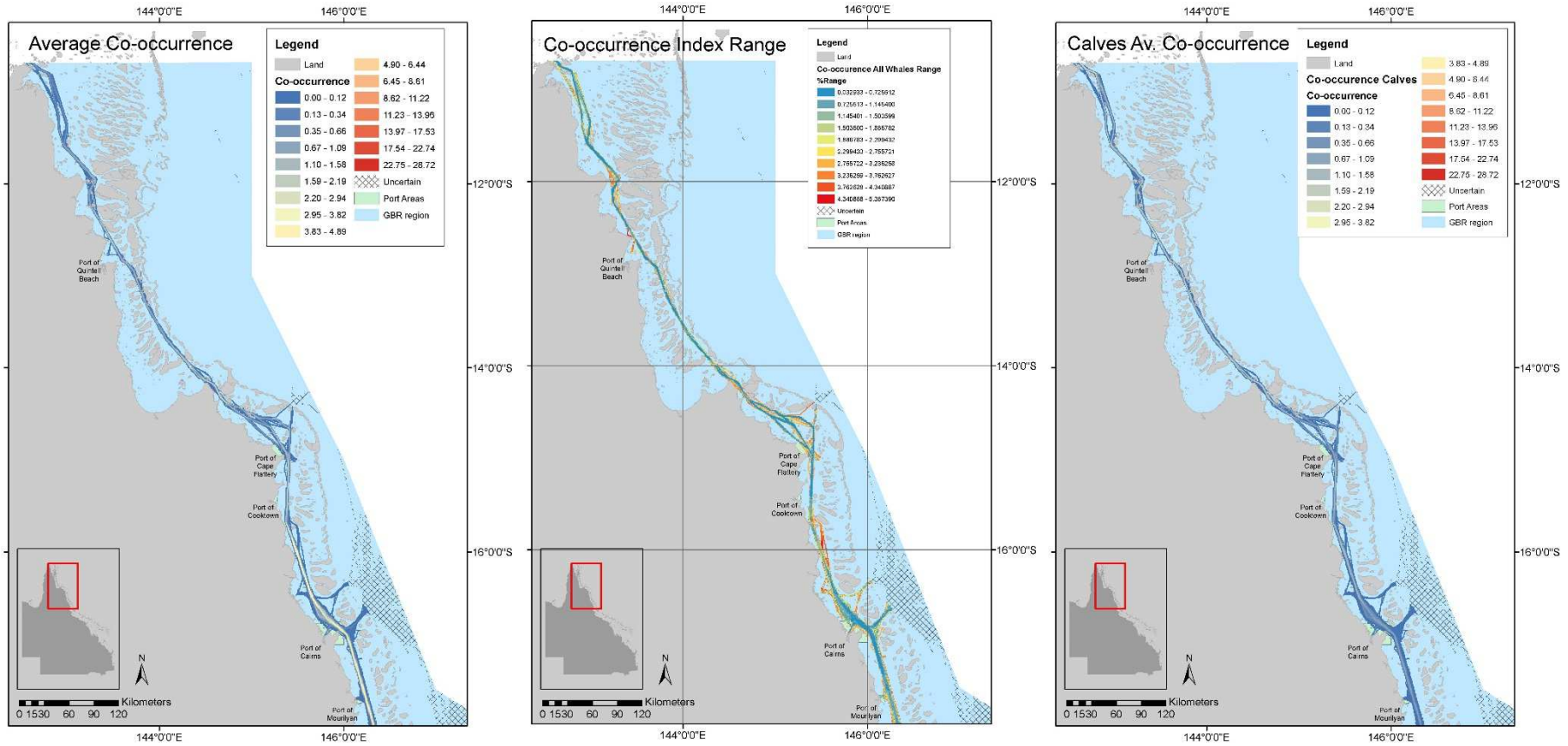


Figure 31: Co-occurrence risk all whale groups (left) and its standardised range (middle) and co-occurrence for calf groups (right), for all vessel type over the Northern extent

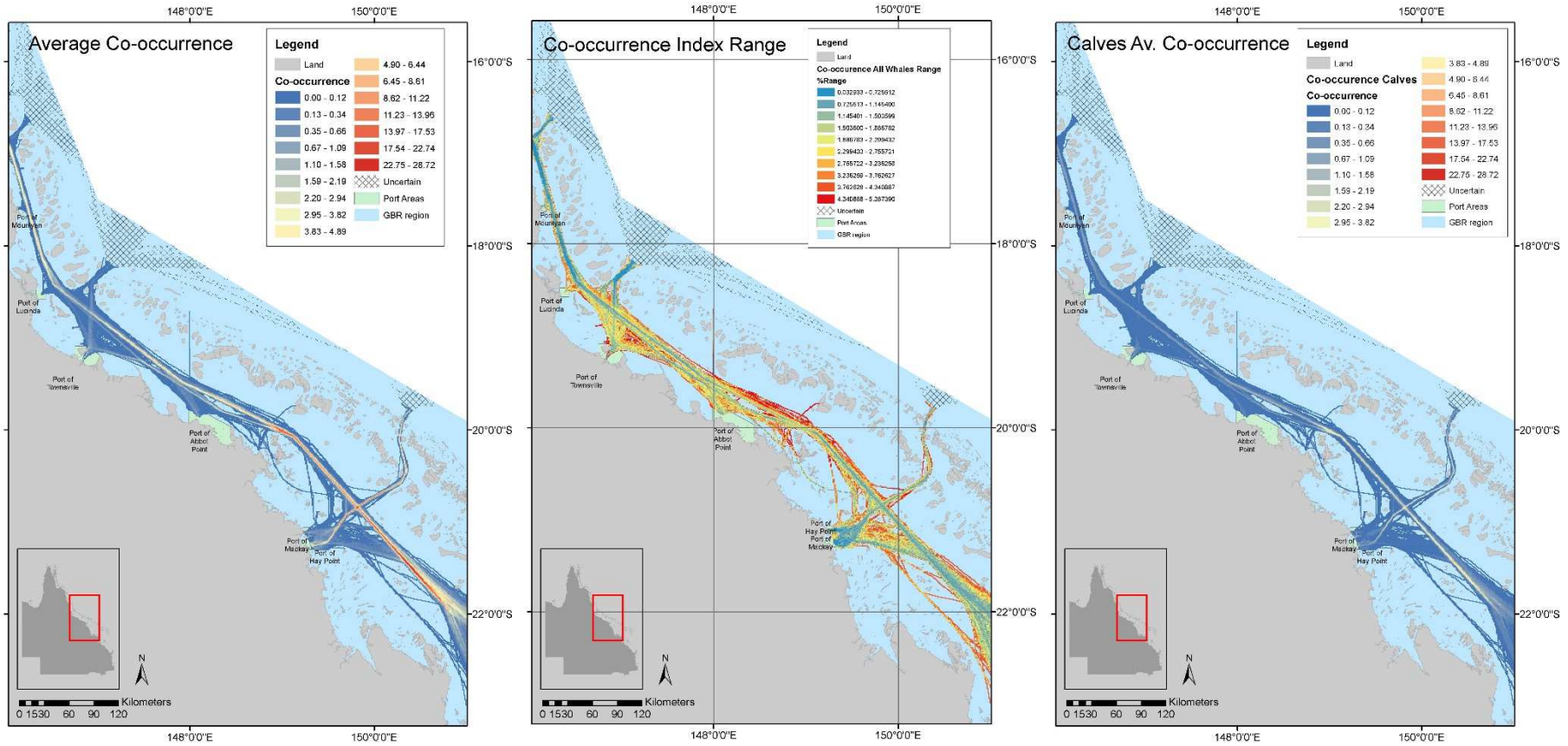


Figure 32: Co-occurrence risk all whale groups (left) and its standardised range (middle) and co-occurrence for calf groups (right), for all vessel type over the central extent

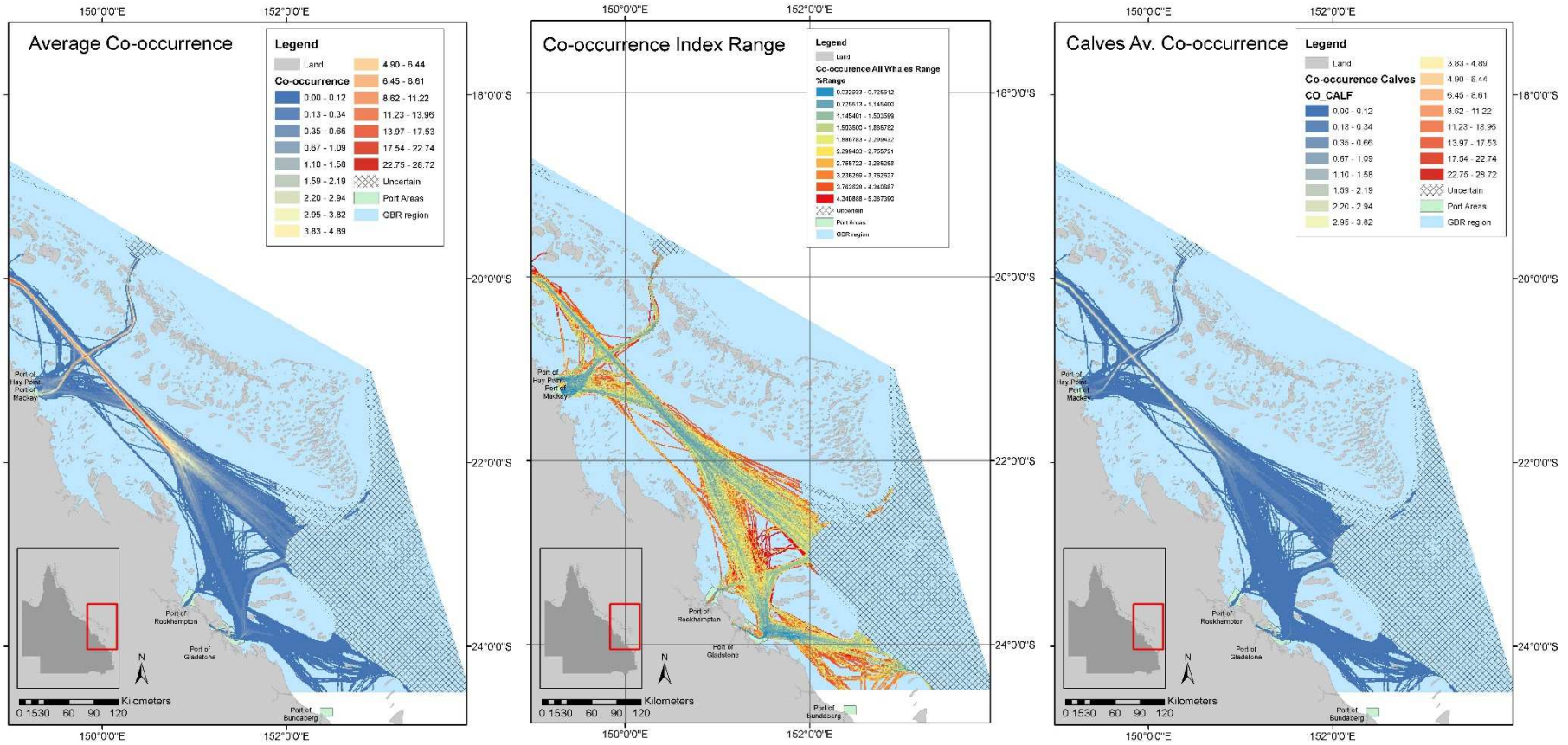


Figure 33: Co-occurrence risk all whale groups (left) and its standardised range (middle) and co-occurrence for calf groups (right), for all vessel type over the Southern extent

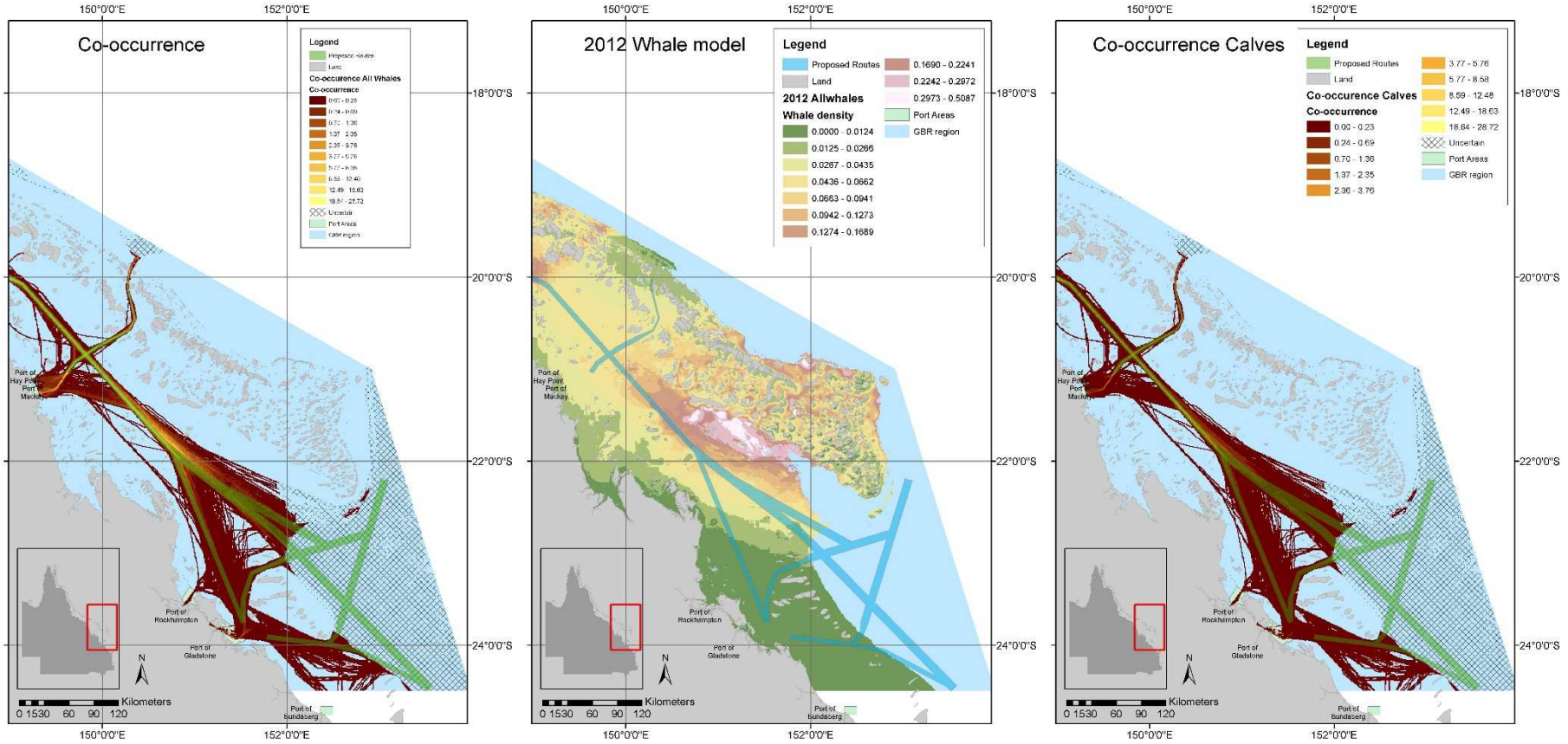


Figure 34: The IMO adopted shipping routes overlayed on the Co-occurrence risk all whale groups (left) and the 2012 whale density predictions (middle) and co-occurrence for calf groups (right), for all vessel type over the Southern extent

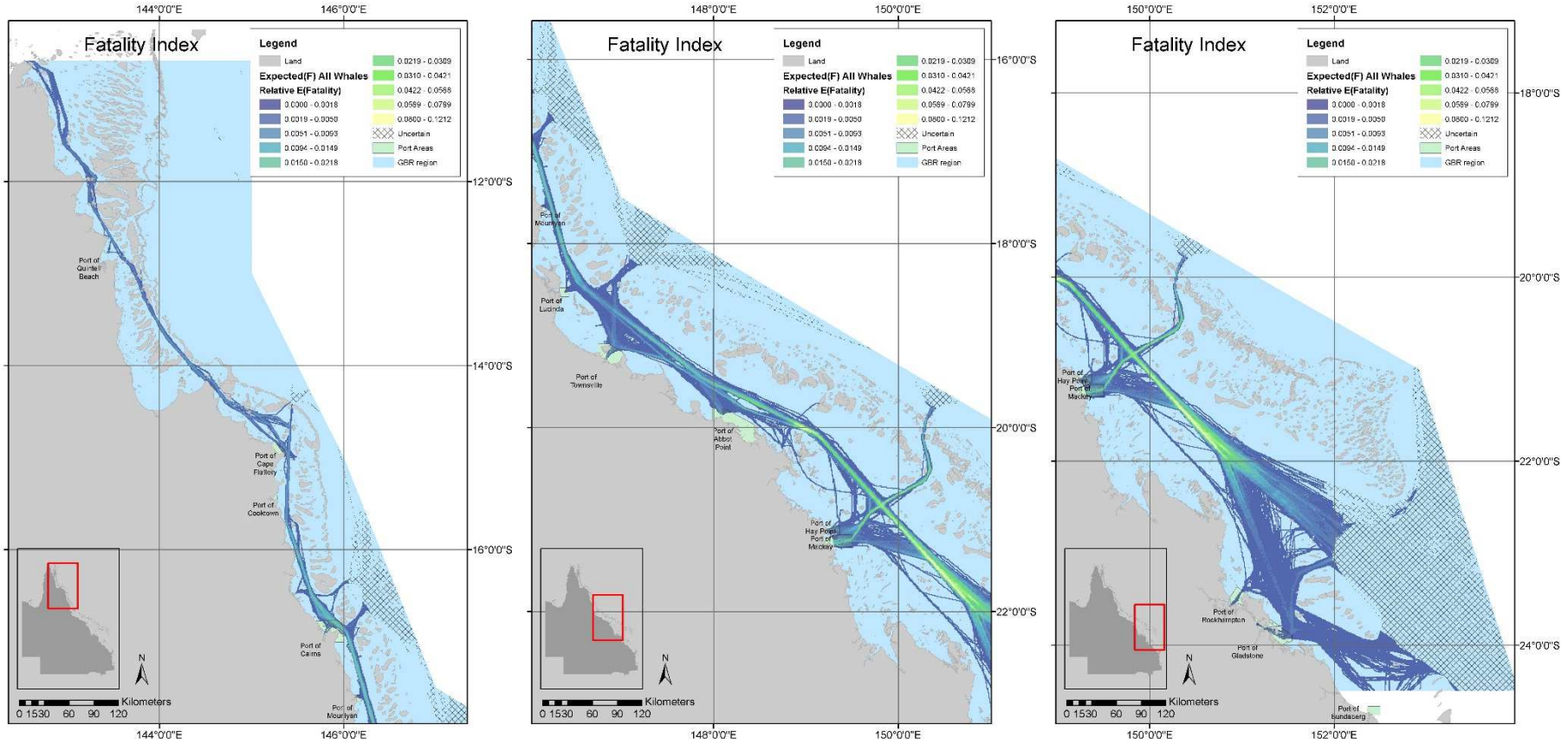


Figure 35: Relative probability of fatality for the 2014 year analysis

4.5.5 Uncertainty

The final risk values are the combination of whale abundance and shipping density, and both of these components contain a degree of uncertainty. Quantifying uncertainty is important not only to indicate how much trust should be given to the overall results, but also when it comes to using the application in informing spatial decisions to manage risk. For example, if considering two route locations and one has a slightly lower relative risk, it would seem preferable to recommend that route. However, if we consider uncertainty and the lower risk route has very large uncertainty and the slightly higher relative risk route has very little uncertainty it then may be preferable to choose the slightly higher risk route.

The main sources of uncertainty include:

- Inter-annual variation in spatial distribution
    - There is little large scale information available about inter-annual variation in distribution within the GBRWHA but there are some small scale projects that could be used to potentially assess this including long term data sets in places such as the WhitSunday Islands. How any fine scale variation in distribution from year to year could be translated into a large scale would be challenging. Based on limited satellite tagging of humpbacks (Gales et al. 2010) it does appear that most of the northward migrating whales from South East Queensland progress into the GBRWHA.
    - The best data set for exploring this issue would be the 2012 and 2014 aerial surveys data and specifically the area that was surveyed during both periods where this could be assessed to provide an indication of inter-year spatial variation.
    - Another source of related uncertainty is that as the population is recovering well, are we going to see increasing whale densities in the same areas or are we going to see whale densities remain constant but with an increased spatial coverage. It is possible that the West coast population (IWC Stock D) could be capable of informing this question due to the greater population size and anecdotal reports of humpback whales sighted in areas further north than the main identified breeding area (*pers. comm.* R. Groom). The populations in Hawaii may also be another potential source of information.
    - Uncertainty of spatial distribution is likely to have a larger impact than uncertainty around numbers as we have considerably better data for the latter. For example if humpback location is more of a random process than driven by habitat or other covariates, then it won't be easy to measure and model it.
- Inter-annual variation in total numbers
    - The population of humpbacks on the East coast of Australia is growing at an extraordinarily rate of 11% per annum (Noad et al. 2011). While the rate of increase in the population is well documented, with land-based population surveys undertaken every few years, there is a degree of uncertainty regarding the proportion of the population that undertakes the migration every year from the Antarctic waters and whether there is inter-annual variation in migration whale numbers. Nevertheless, if the rate of population increase continues then inter-annual variation in migration numbers may not necessarily have a large impact on the risk assessment due to the main factor being increasing numbers of whales on the breeding ground in the GBRWHA.
- Intra-season variation in numbers and spatial distribution
    - It is well known that there is variability within seasons in terms of both the number and distribution of humpback whales. This is primarily a function of the GBRWHA being a breeding and mating area with only temporary residency times and whales move into the area, mate and calve and move out again over an approximately 3-4

month period. This means that relative risk will vary through the season as densities and distribution varies. However, while this issue is well known is poorly described as there have been no large scale intra season surveys from which to assess this.
    - The aerial survey data that we have used in developing spatial models was aimed to coincide with the peak of numbers within the GBRWHA and so our model is likely to represent that period of highest risk with periods before and after the survey likely to have lower whale densities.
    - Ideally, it would be desirable to undertaken multiple surveys of the same region/s within a season to investigate variability in both numbers and distribution.
- Whale model uncertainty
    - All indications from overseas applications of this type of modelling approach are that the model standard errors will be swamped by the natural annual temporal variability. Notwithstanding this, it is important to include model uncertainty as it can be a significant contributor to overall uncertainty.
    - We have standard errors and confidence intervals on the model-based whale density predictions. These can also be incorporated into the risk calculations, at the most simplest by producing results using the confidence intervals and reporting the corresponding risk maps.
- Inter-annual variation in vessel movements
    - Given the availability of shipping data this should be easy to quantify and given the shipping routes are pre-defined this is unlikely to be a large issue.
    - One approach would be to run the various years of shipping data against the two whale years and look at variation in the final risk estimates.

Overall, our approach was to uncertainty was to include some indicative estimates to explore model sensitivity. We ran each combination of shipping data year (2012, 2013, and 2014) against each whale model year (2012, 2014) at each confidence limit. Then in each cell we report the minimum and maximum risk.

4.5.6 Assumptions

A number of assumptions were made in the risk calculations. Table 9 provides a summary of the main assumptions.

Table 9: Main assumptions

	Assumption	Comment
AIS Data	Only vessels >80 m in length in the analysis	This decision was based on experience overseas with large whales. The analysis can easily be re-run at a lower cut-off and the results compared.
	Invalid or uncertain data removed	This will reduce the calculated ship strike risk values. However, since at this stage we are dealing with relative risk as long as the missing, invalid or corrupt data is randomly distributed with respect to our variables of interest (i.e., spatially, vessel speed and vessel type) no bias will be

		introduced
Whale Model	There is no inter-season variation	We know this not to be true and is one of the aspects of uncertainty that needs further work.
	Various model and distributional assumptions	See section 3.2. Many of these are standard statistical assumptions when (1) doing distance analysis and (2) fitting GAMs
Co-occurrence Risk	That risk is proportional to the number of ships (actually the distance covered) and whales in an area and that they equally contribute to risk	This seems a fairly reasonable assumption to make.
	That the relationship between risk and whale numbers in the grid cell is linear	Again this assumption seems reasonable. One reason it may not hold is if there was heterogeneity in the whale population with regard to strike risk that is not considered. For example, mother-calves have a much higher risk than non-calf groups. Then adding whales to a cell to double the number of whales would not necessarily double the ship strike risk as it would depend on the makeup of the existing whales and the added whales.
	That the relationship between risk and vessel numbers in the grid cell is linear	Again this seems reasonable, but again if there are vessel differences that are not considered this may not be true.
	That whale distribution is independent of vessel traffic	An example where this would not be the case is if in general due to noise levels whales avoided certain areas.

5. Discussion

5.1 Humpback habitat suitability mapping & density estimation

The aerial surveys were very successful in describing the distribution and density of whales within the survey areas. Although encounter rates were higher in 2012 than in 2014, the overall densities of all groups and groups with calves were not significantly different. Surveys were undertaken from July through to September which coincides with the highest densities of whales in the region. Owing to the relatively large number of humpback whale sightings, the MRDS approach was successful in allowing us to maximise the efficient use of the data and to develop statistically robust models of humpback distribution and density. One of the limitations of the approach was that while relatively large areas of the GBRWHA were covered during the aerial surveys, it was impossible to cover the entire GBRWHA and so it was therefore necessary to use model-based approaches to investigate and quantify whale distribution outside the surveys areas.

The best GAM, or density surface model, describing the distribution of humpback whale densities included the physiographic and environmental covariates bathymetry, sea surface temperature and sea surface height anomalies. While a wide range of potential covariates were explored, only these three were influential. The MAXent modelling approach by Smith et al. (2012) identified the three most informative environmental predictors for humpback habitat suitability as sea surface temperature, depth and distance from the coast. The two approaches utilised different data sets and methods but provided similar results providing further confidence in the results. We are unable to explain why sea surface anomaly is likely to be a significant variable. There was no evidence of any significant correlations between the range of spatial and environmental covariates as any correlations were small enough to ignore.

The general results predicted that higher densities of humpbacks were more likely to be found in shallow water (e.g., ~20-60 m deep), in waters of a sea surface temperature of ~21-23°C and with a sea surface height anomaly of ~0.05 m. These are consistent with our general understanding of habitat requirements for breeding and mating humpback whales in Australia.

Owing to only a small amount of survey effort in bathymetric values of 90 m and deeper, no density predictions were made for waters deeper than 90 m. This is a potential limitation of this analysis and would require addition aerial surveys to rectify. For the purposes of this application, we excluded any waters deeper than 90 m which lead to limited data being available for the Bunker/Capricorn reef systems in the southern part of the GBRWHA. It will be important to address this limitation considering this area has previously been identified as an important area for humpback whales, most likely as a concentrated migration route (Smith et al. 2012). An additional issue is that there has been no survey effort to the far north of the GBR and while it is expected that whale density will be low based on anecdotal reports, the model would be improved by aerial survey data from that region, although this is likely less of an issue than the limited data from the southern GBR.

Overall, there was good consistency between the aerial survey and spatial modelling predictions. In addition, there was relatively good agreement between the previous habitat suitability modelling by Smith et al. (2012) and the distribution of densities of humpback whales extrapolated throughout the GBR using our density surface model. Based on these assessments, we have good confidence that the model outputs have been successful in describing relative whale densities and in identifying areas of high whale density. This provided us with an excellent understanding of whale distribution and density for comparison to vessel traffic.

5.2 AIS shipping data

We were able to access high quality AIS vessel tracking data from AMSA available through the CTS data series starting in June 2012. AMSA were extremely helpful in making the data available for this research project and aiding with interpretation and advice. This provided us with a three year data set for the months of July, August and September which overlapped with the aerial survey data from 2012 and 2014. While we have only utilised the most relevant parts of the available data, there is scope to undertake additional analysis and investigate other aspects of the shipping data that would provide an improved insight into vessel behaviour and could be used to further refine aspects of the modelling. In addition, although not an issue for this specific study due to the size of humpback whales, it should be noted that while the AIS data set is suitable for describing the behaviour of large vessels which have mandatory reporting requirements, smaller vessels only have voluntary reporting and therefore it is unlikely to be comprehensive for these types of vessels.

There was a considerable amount of work involved in validating, filtering and processing the data before it was in a form suitable for the calculation of relative risk. The overarching feature was that AIS data is time based and it was necessary to convert it to distance based data so it could be compared to humpback whale spatial data to allow for risk-based assessments to be undertaken.

Overall, we included over 2.3 million data points post-filtering in analysis from three months of interest.

During data filtering it was necessary to make some decisions about data criteria. Most were straight forward but perhaps potentially the most influential one was filtering on vessels ≥ 80 m. The main reason for this was that, as a general rule, larger vessels pose the highest risk of causing death to humpbacks from ship strike. This assumption was based on overseas experience and follows previous risk-assessment approaches for ship strike (e.g., Redfern et al. 2013 analysed vessels > 100m). By excluding smaller vessels, we are potentially negatively biasing our estimates of risk and therefore estimates should be considered as minimums. However, if vessel movements for these smaller vessels follow the same general distribution and density of the larger vessels, then this assumption will not cause a bias in relative risk.

There was a slight increase in the amount of shipping traffic over the three years period but no discernible difference by month. The vast majority of vessels in the filtered data set were cargo vessels, followed distantly by tankers and a small number of passenger vessels. This partly reflects our assumption of only using vessels over 80 m in length as the bulk of passenger vessels will be shorter than this. Overall, the final data set can be summarised as being comprised of vessels with a length ≥ 80 m representing cargo, tanker and passenger types within the GBR during the winter humpback breeding season.

In future it would be beneficial to get NAV_STATUS from the AIS data to aid in the filtering of vessels at anchor.

5.3 Ship strike risk

The aim of this research was to quantify relative, rather than absolute risk, and therefore, it is necessary to have a suitable metric. We implemented two different metrics comprising of an existing risk metric based on the idea of co-occurrence of whales and vessels, and a more complex metric using a probabilistic framework to provide a relative index of the expected number of ship strike mortalities.

The first metric implemented was the simple measure of co-occurrence, which assuming other variables are constant spatially, should be proportional to risk. The measure of co-occurrence in a particular grid cell was simply taken to be the total distance traversed in a cell by ships multiplied by the number of whales in the cell.

We also developed and tested the more advanced relative index of the number of expected whale fatalities. This framework can incorporate the effect of different vessel characteristics (e.g., speed, beam, and draught) and conceptually includes terms related to time at surface and avoidance. As new data becomes available (e.g., dive data), the approach can easily be extended to incorporate other aspects such as whale sub-groups with differing risk profiles (e.g., mother with calves versus adult differences in diving and vessel avoidance).

When considering overall risk of ship strike, it was evident that cargo vessels provided the single largest contribution (Figure 33). This is not to say individual cargo vessels are intrinsically pose more of a risk, as the co-occurrence measure does not take vessel characteristics into account (i.e., speed, beam, etc.), the difference reflects the large number of cargo vessels relative to tanker and passenger (Figure 21). This is demonstrated when risk is standardised by km traversed by vessels of each type to give risk per vessel km (Figure 33), the relative risk of a typical single cargo vessel is comparable with a single tanker. The passenger vessels risk per vessel km (Figure 33) is higher relative to cargo and tanker. Since no speed or other vessel characteristic is incorporated in the co-occurrence measure, this could only be due to a different spatial distribution of where passenger vessels travel. However, in this case, we believe it could just reflect the small number of passenger

vessels (n=15) in the analysis giving a variable result. In terms of overall risk, which is what is important, cargo and tanker vessels are of the most concern.

When considering relative risk for each whale group type (i.e., groups with or without calves), once the risk is standardised for the number of total animals of each group type to give risk per whale, then there was a slightly higher co-occurrence observed for groups with calves. Given that the co-occurrence model just reflects the number of whales and vessels that are spatially co-existing, the absolute number of each type will directly influence relative risk. Further, data and analysis would be needed to discern if in general groups with calves are any more susceptible to ship strike risk than non-calf groups, as the current co-occurrence index assumes group types are equally likely to be struck (as it does not consider differing risk of ship strike due to differences in whale behaviour).

Based on the co-occurrence maps it appears the area of greatest relative risk is two areas located approximately 120km to the North and 120km South of Mackay. After examining the whale habitat models it is clear these correspond to where shipping traverses two higher predicted whale density areas.

Our approach provides an indication of variability in the modelling. Perhaps unsurprisingly, overall, the modelling indicates that the major shipping lanes appear to be less variable with relatively fixed and discrete edges. By comparison, outside the main lanes, the edges of routes are more variable but in general these also have fewer vessels transiting through them. This makes sense as our approach will capture vessel temporal variability in space much better than it will variability in whale distribution (as we only had two whale seasons to assess variability).

We developed a relative index of the expected number of fatalities as a proof of concept and to investigate if this is a feasible approach. It is a useful way of exploring the modelling framework by producing a metric that may be useful in reviewing possible impacts and informing possible management. The outcomes from this modelling was similar to that of the co-occurrence modelling. At this point, the index it is simply a relative metric across the study area useful to comparing relative risk and cannot and should not be inferred as an estimation of actual mortality. While this step is potentially possible, it would require the considerable data specific to this issue to be collected before the leap between relative and actual mortality rates could be made. Notwithstanding this major caveat, it is a powerful tool that can be useful in:

- easily incorporating vessel characteristics (e.g., speed, length, width) and can aid with making improved comparisons;
- investigating expected risk reduction outcomes from management decision such as speed restrictions;
- incorporating different whale strike risks (e.g., mother calf vs adult groups); and
- potentially having less bias at higher ship densities than co-occurrence (See section 3.4.2)

Overall, the quantification of relative risk over this large spatial scale has been successful and is useful in identifying areas of high co-occurrence. The analysis shows that the areas of highest relative risk coincide with offshore areas around the two major ports on the Queensland coast spanning the offshore area between the Whitsundays to south of Mackay near Shoalwater Bay. One limitation of this risk assessment is that there are no results for the offshore areas to the south of Gladstone due to a lack of whale survey data in deeper waters.

5.4 Modelling uncertainty of ship strike

The final relative risk values are the combination of whale and shipping density, and both of these components contain a degree of uncertainty. Quantifying uncertainty is important not only to indicate how much trust should be given to the overall results, but also when it comes to using the application in informing spatial decisions to manage risk. The largest source of uncertainty is likely to be related to inter-annual variation in the spatial distribution of whales as there is little information

data to inform this but without additional surveys replicating the coverage of previous ones, this will be difficult to quantify. Other sources of uncertainty include inter-annual variability in whale abundance and vessel behaviour, intra-season variation in numbers and spatial distribution of whales, and inherent uncertainty of the model itself although with respect to this latter issue, all indications from overseas applications of this type of modelling approach is that the model standard errors will be swamped by the natural annual temporal variability.

5.5 Future extensions

We have identified a range of possible extensions to the various modelling approaches throughout the document. Specifically, there were some recommendations about potential ways to increase the data that underpins the model including:

- Dedicated studies of whale behaviour within the areas of high co-occurrence. This could be undertaken using DTAGs and/or ZTAGs to understand dive and movement behaviour in the presence and absence of vessels. The DTag dive depth data could be used to determine an estimate of absolute probability of a fatal ship strike;
- Investigation of the use of AIS data for the monitoring of vessels < 80 m in length and whether it is possible to use this data for reliable analysis of passenger vessel movements which was poorly covered in this project due to a lack of data. A greater sample size of vessels < 80m will improve relative risk estimates when comparing among different vessel types; and
- Aerial surveys expanding the range into (1) waters > 90 m deep for which we think may be important whale habitat; (2) exploration of the northern range of the GBR which has had very little survey effort; and (3) resurveying areas that have previously been surveyed to provide estimates of inter-annual variability which will all improve model predictions and estimates of uncertainty

5.6 Mitigating ship strike risk

The major output of this approach is that it has highlighted areas were relative ship strike risk and/or the expected number of fatalities are high. While the true relationship between relative and actual risk remains unknown, these data provide the best source of information to aid in the identification of potential hotspots of high interactions. These are the areas that will require further consideration both with respect to targeted research and also the exploration of whether management action has the potential to improve relative risk.

There is no silver bullet for the mitigation of ship strike risk, as whenever whales and vessels occupy the same space there is always risk of a negative interaction. However, now that we have identified areas that we believe represent higher risk, we can seek to better understand the actual nature of the interaction in those areas with a view to proposing any number of a suite of potential management actions. These could range from ‘no action required’ through to active management, including options such as zones that could require speed reductions, requirements for increased observation of marine mammals in the path of vessels by bridge crew and modification of vessel routes to avoid areas of higher whale density.

6. Acknowledgements

We are extremely grateful to many people and organisations who have aided in the development of this project. Specifically to:

- our primary funders, the Australian Marine Mammal Centre Grant Programme and the International Fund for Animal Welfare Oceania Office;
- CSIRO, Murdoch University and Blue Planet Marine for supporting us including contributions of our time in undertaking this project;
- Our colleagues and collaborators Jessica Redfern and TJ Moore of NOAA who have been instrumental in supporting and sharing their time, expertise and Code;
- The aerial charter company Observair and pilot Brad Welch for professional and experienced flying and the aerial survey team in 2014 who did an excellent job in collecting high quality data from the GBR, including team leader Louise Bennett, Maria Jedjenso, Shannon McKay, Kylie Mackenzie and Amy James;
- IFAW including Isabel McCrea, Matt Collis and Sharon Livermore;
- Mark Read and Andrew Simmons at GBRMPA for support, advice and access to data;
- Edward King (CSIRO) for advice on environmental covariates; and
- AMSA for the provision of AIS vessel data and helpful advice especially Ross Henderson.

7. References

Andriolo, A., Martins, C. C. A., Engel, M. H., Pizzorno, J. L., Más-Rosa, S., Freitas, A. C., ... & Kinas, P. G. (2006). The first aerial survey to estimate abundance of humpback whales (*Megaptera movaeangliae*) in the breeding ground off Brazil (Breeding Stock A). Journal of Cetacean Research and Management, 8(3), 307.

Baird, R. W., Ligon, A. D., & Hooker, S. K. (2000). Sub-surface and night-time behavior of humpback whales off Maui, Hawaii: a preliminary report. Report prepared under Contract, (40ABNC050729).

Braemar Seascope (2013) North Queensland ship traffic growth study. Supplementary Report. 22 March 2013. 71 p.

Bravington MV, Hedley SL (2010) Antarctic minke whale abundance from the SPLINTR model: some 'reference' dataset results and 'preferred' estimates from the second and third circumpolar IWDC/SOWER surveys. Paper SC/62/IA12 presented to the IWC Scientific Committee, June 2010 (unpublished). 14pp.

Burt ML, Borchers DL, Jenkins KJ, Marques TA (2014) Using mark-recapture distance sampling methods on line transect surveys. Methods in Ecology and Evolution 5(11): 1180-1191.

DEH (2005) Humpback Whale Recovery Plan 2005 - 2010. Department of the Environment and Heritage, Commonwealth of Australia, Canberra

ESRI (2012) ArcGIS Desktop: Release 10.1. Redlands, CA: Environmental Systems Research Institute.

Forestell, PH, Kaufman GD, Chaloupka M (2003) Migratory Characteristics of Humpback Whales (*Megaptera novaeangliae*) in Hervey Bay and the Whitsunday Islands, Queensland, Australia: 1993-1999. Final Report to Environmental Protections Agency, Queensland Parks and Wildlife Service, Brisbane, Australia. , pp 57

GBRMPA (Great Barrier Reef Marine Park Authority) (2013) Ports and shipping information sheet: shipping – challenges for the Great Barrier Reef. Accessed 6 June 2014. Available at: http://www.gbrmpa.gov.au/__data/assets/pdf_file/0020/28811/Ports-and-Shipping-Info-Sheet-4-Shipping-May-2013-update.pdf

Hedley SL, Buckland ST (2004) Spatial models for line transect sampling. Journal of Agricultural Biological and Environmental Statistics 9(2): 181-199.

Integrated Marine Observing System (IMOS) (2015a) SRS Satellite – SST L3S – 1 day composite – day and night time, https://imos.aodn.org.au/imos123/home, accessed 1 April 2015.

Integrated Marine Observing System (IMOS) (2015b) Gridded Sea Level Anomaly-Delayed Mode, https://imos.aodn.org.au/imos123/home, accessed 1 April 2015.

Integrated Marine Observing System (IMOS) (2015c). SRS Satellite-OC MODIS – 1 day – Chlorophyll a concentration algorithm, https://imos.aodn.org.au/imos123/home, accessed 1 April 2015.

Jørgensen B (1987) Exponential dispersion models. Journal of the Royal Statistical Society Series B 49: 127-162.

Laake JL, Borchers DL (2004) Methods for incomplete detection at distance zero. Advanced Distance Sampling.

Laake JL, Borchers D, Thomas L, Miller D, Bishop J (2015) MRDS: Mark-Recapture Distance Sampling (MRDS): R package.

Marques FFC, Buckland ST (2003) Incorporating covariates into standard line transect analyses. Biometrics 59(4): 924-935.

Marques FFC, Buckland ST (2004) Covariate models for the detection function. Advanced Distance Sampling. S. T. Buckland, D. R. Anderson, K. P. Burnham et al. Oxford, Oxford University Press.

Miller DL, Burt ML, Rexstad EA, Thomas L (2013) Spatial models for distance sampling data: recent developments and future directions. Methods in Ecology and Evolution 4(11): 1001-1010.

Noad MJ, Paton D, Rekdahl M (2009) Hay Point Whale Survey 2009. Pages 20. Draft report for Sinclair Knight Merz.

Noad MJ, Dunlop RA, Paton DA, Kniest H (2011) Abundance estimates of the east Australian humpback whale population: 2010 survey and update, Paper submitted for consideration by the IWC Scientific Committee. SC/63/SH22, pp 12

North-East Shipping Management Group (NESMG) (2014) North-East Shipping Management Plan. October 2014. 132 p.

Paton, D & Kniest, E 2011, 'Population growth of Australian east coast humpback whales, observed from Cape Byron, 1998 to 2004', The Journal of Cetacean Research and Management, no. Special issue 3, pp. 261-268.

Pollock KH, Marsh HD, Lawler IR, Alldredge MW (2006) Estimating animal abundance in heterogeneous environments: an application to aerial surveys for dugongs. Journal of Wildlife Management 70: 255-262

R Core Development Team (2015) R: A language and environment for statistical computing. Austria, R Foundation for Statistical Computing.

Redfern JV, Mckenna MF, Moore TJ, Calambokidis J, Deangelis ML, Becker EA, Barlow J, Forney KA, Fiedler PC, Chivers SJ (2013) Assessing the risk of ships striking large whales in marine spatial planning. Conservation Biology 27: 292-302

Silber, G. K., Slutsky, J., & Bettridge, S. (2010). Hydrodynamics of a ship/whale collision. Journal of Experimental Marine Biology and Ecology, 391(1), 10-19.

Skaug HJ (2006) Markov modulated Poisson processes for clustered line transect data. Environmental and Ecological Statistics. 13: 199-211.

Smith JN, Hedley S (in prep) Breeding grounds of humpback whales in the Great Barrier Reef World Heritage Area: validation of a predictive spatial habitat model

Smith JN, Noad MJ, Grantham HS, Paton D (2012) Identification of humpback whale breeding and calving habitat in the Great Barrier Reef. Marine Ecology Progress Series 447: 259-272

Vanderlaan ASM, Taggart CT (2006) Vessel collisions with whales: the probability of lethal injury based on vessel speed. Marine Mammal Science 23: 144-156.

Williams R, Hedley S, Branch TA, Bravington M, Zerbini AN, Findlay K (2011) Chilean Blue Whales as a case study to illustrate methods to estimate abundance and evaluate conservation status of rare species. Conservation Biology 25: 526-535.

Wood SN (2006) Generalized Additive Models: An Introduction with R. USA, Chapman and Hall.

Appendix 1 Documentation of model-selection results for MRDS detection function fitting

Table A1.1 Hazard rate MRDS models fitted to the sighting data, with a left truncation distance of around 200 m, and a right truncation of 4 km. BBS = Raw Beaufort Sea State, size = raw group size estimate, bbs.0.1 = categorise Beaufort Sea State, Cloud.cover = cloud cover in octas. ΔAIC values are relative to the lowest model AIC, given in the first row.

DS Model	MR Model	AIC	ΔAIC
1	BSS+distance+size	2016.659	0
1	distance+size	2017.326	0.667128
1	distance+size	2017.326	0.667128
size	BSS+distance+size	2017.903	1.243559
1	bss.0.1+distance+size	2017.928	1.268494
BSS	BSS+distance+size	2018.204	1.544638
Cloud.cover	BSS+distance+size	2018.469	1.809469
size	distance+size	2018.57	1.910686
size	distance+size	2018.57	1.910686
1	BSS+Cloud.cover+distance+size	2018.648	1.988769
BSS	distance+size	2018.871	2.211766
BSS	distance+size	2018.871	2.211766
1	Cloud.cover+distance+size	2019.075	2.415673
1	Cloud.cover+distance+size	2019.075	2.415673
Cloud.cover	distance+size	2019.136	2.476596
Cloud.cover	distance+size	2019.136	2.476596
size	bss.0.1+distance+size	2019.171	2.512053
BSS	bss.0.1+distance+size	2019.472	2.813132
BSS+size	BSS+distance+size	2019.707	3.047436
Cloud.cover+size	BSS+distance+size	2019.726	3.066899

Table A1.2 Half-Normal MRDS models fitted to the sighting data, with a left truncation distance of around 200 m, and a right truncation of 4 km. BBS = Raw Beaufort Sea State, size = raw group size estimate, bbs.0.1 = categorise Beaufort Sea State, Cloud.cover = cloud cover in octas. ΔAIC values are relative to the lowest model AIC, given in the first row.

DS Model	MR Model	AIC	ΔAIC
1	BSS+distance+size	2020.729	0
1	distance+size	2021.396	0.667128
1	distance+size	2021.396	0.667128
size	BSS+distance+size	2021.716	0.986575
1	bss.0.1+distance+size	2021.998	1.268494
BSS	BSS+distance+size	2022.073	1.343992
size	distance+size	2022.383	1.653703
size	distance+size	2022.383	1.653703
1	BSS+Cloud.cover+distance+size	2022.718	1.988769
Cloud.cover	BSS+distance+size	2022.724	1.995098
BSS	distance+size	2022.74	2.01112
BSS	distance+size	2022.74	2.01112
size	bss.0.1+distance+size	2022.984	2.255069
1	Cloud.cover+distance+size	2023.145	2.415673
1	Cloud.cover+distance+size	2023.145	2.415673
BSS+size	BSS+distance+size	2023.162	2.432424
BSS	bss.0.1+distance+size	2023.342	2.612486
Cloud.cover	distance+size	2023.392	2.662225
Cloud.cover	distance+size	2023.392	2.662225
size	BSS+Cloud.cover+distance+size	2023.705	2.975344

Details of best detection function fit as outlined in Table A1.1 above:
Distance sampling analysis object

Summary for io.fi object

Number of observations:	561
Number seen by primary:	488
Number seen by secondary:	396
Number seen by both:	323
AIC:	1028.964

Conditional detection function parameters:

	Estimate	SE
(Intercept)	1.3165169	0.3173286
BSS	-0.4994220	0.1305199
Distance	-0.5100542	0.1038039
Size	0.7222984	0.1347244

	Estimate	SE	CV
Average primary p(0)	0.8396466	0.024371004	0.02902531
Average secondary p(0)	0.8396874	0.024364279	0.02901589
Average combined p(0)	0.9678823	0.009840597	0.01016714

Summary for ds object

Number of observations:	561
Distance range:	0.2 - 4
AIC:	1297.803

Detection function: Hazard-rate key function

Detection function parameters

Scale Coefficients:

	Estimate	SE
(Intercept)	0.7114042	0.05885498

Shape parameters:

	Estimate	SE
(Intercept)	1.147507	0.1293917

	Estimate	SE	CV
Average p	0.5713126	0.02361902	0.04134167

Summary for io object

Total AIC value : 2326.767

	Estimate	SE	CV
Average p	0.5529633	0.0235416	0.04257352
N in covered region	1014.5338364	51.8356906	0.05109311

Appendix 2 Results for 'space' only GAMs for describing the distributions of densities of humpback whales, and humpback whales in groups accompanying calves, for both surveys

Note: Space only GAMs indicates those which only contained spatial coordinates (in this example, latitude and longitude transformed by an Albers Equal Area projection).

'Space' only GAMs describing the distribution of densities of humpback whales during the *2012* survey.

```
Family: Tweedie(1.4)
Link function: log

Formula:
total.total.n ~ s(Easting, Northing, k = 30) + offset(log.offset)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.8319     0.1027  -37.31   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                     edf Ref.df     F  p-value
s(Easting,Northing) 14.32  18.72 5.198 1.25e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.405   Deviance explained = 32.9%
-REML = 419.42  Scale est. = 2.5968    n = 256
```

'Space' only GAMs describing the distribution of densities of humpback whales during the *2014* survey.

```
Family: Tweedie(1.4)
Link function: log

Formula:
total.total.n ~ s(Easting, Northing, k = 30) + offset(log.offset)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.3021     0.1047  -41.08   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                     edf Ref.df     F  p-value
s(Easting,Northing) 11.72  15.81 7.339 6.87e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.278   Deviance explained = 31.9%
-REML = 433.85  Scale est. = 2.217     n = 323
```

‘Space’ only GAMs describing the distribution of densities of humpback whales in groups containing calves during the *2012* survey.

```
Family: Tweedie(1.1)
Link function: log

Formula:
total.n.with.calves ~ s(Easting, Northing, k = 30) + offset(log.offset)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.8402     0.1533  -31.57   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                      edf Ref.df     F p-value
s(Easting,Northing) 9.836  13.26 2.095  0.0143 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.255   Deviance explained = 19.4%
-REML = 186.38  Scale est. = 2.5852    n = 256
```

‘Space’ only GAMs describing the distribution of densities of humpback whales in groups containing calves during the *2014* survey.

```
Family: Tweedie(1.1)
Link function: log

Formula:
total.n.with.calves ~ s(Easting, Northing, k = 30) + offset(log.offset)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -5.9039     0.2135  -27.65   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                      edf Ref.df    F p-value
s(Easting,Northing) 5.371  7.372 2.42  0.0181 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.0998   Deviance explained = 15.8%
-REML = 112.19  Scale est. = 2.0699    n = 323
```

Appendix 3 Results for best GAMs for describing the distributions of densities of humpback whales, and humpback whales in groups accompanying calves, for both surveys, with physiographic and environmental covariates.

Best GAM describing the distribution of densities of humpback whales (all animals) across both seasons.

```
Family: Tweedie(1.2)
Link function: log

Formula:
total.total.n ~ +te(bathy, SST) + s(SST) + s(SSH) + offset(log.offset)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.02528    0.08203  -49.07   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
               edf Ref.df     F  p-value
te(bathy,SST) 9.058     23 4.183  < 2e-16 ***
s(SST)        2.192      9 0.451  0.00765 **
s(SSH)        2.838      9 3.448 5.96e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.405   Deviance explained = 33.4%
-REML = 821.48  Scale est. = 2.8079    n = 579
```

Best GAM describing the distribution of densities of humpback whales in groups containing calves.

```
Family: Tweedie(1.4)
Link function: log

Formula:
total.n.with.calves ~ te(SSH, bathy) + te(SST, bathy) + offset(log.offset)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -5.4473     0.1651     -33   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Approximate significance of smooth terms:
                edf Ref.df     F p-value
te(SSH,bathy) 4.521  5.322 3.723 0.00209 **
te(SST,bathy) 2.916 20.000 0.510 0.00624 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

R-sq.(adj) =  0.233   Deviance explained = 16.4%
-REML = 343.07  Scale est. = 5.67      n = 579
```

Appendix 4 Additional exploratory analyses of covariate data for spatial modelling of humpback whale aerial survey data from 2012 and 2014

The following provides details of analysis undertaken as part of checking and preparation of aerial survey data to use in spatial modelling.

A4.1 Variations in flight (on-effort) altitude

The recorded flight altitude whilst on effort varied between 900 and 1300 ft for the 2012 survey, and 880 and 1560 ft (267 to 475 m) during the 2014 aerial survey. This could be viewed as substantial variation in effort altitude, which may introduce bias into the detection function and, ultimately, density and abundance estimates (Laake 2008). However, both the mean altitude height (mean weighted by the along-track distance at each altitude) was 999 ft in 2012 and 1003 ft for the 2014 aerial survey were almost exactly the nominated survey altitude of 1000 ft. Therefore there does not seem to be any spatial correlation (viz, along the GBR) in changes to altitude (Figure A4.1); and that it can be assumed that g(0) will not vary from 1 at these ranges of altitudes, then by assuming pooling robustness over the periods of different altitudes, the detection function will still provide an unbiased estimate of densities along the transects.

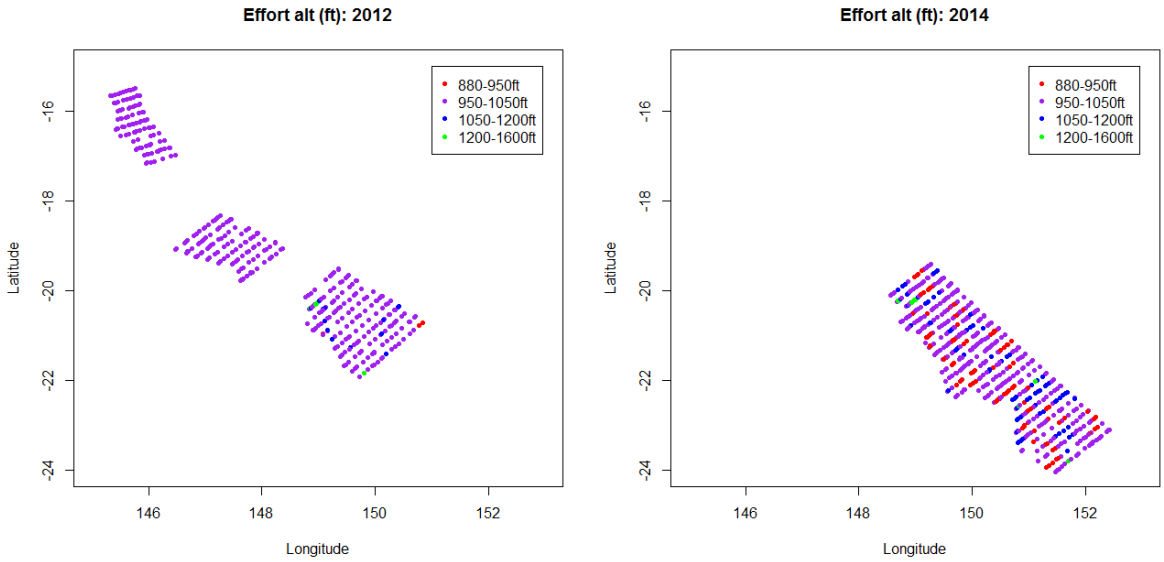


Figure A4.1 Spatial distribution of 'on effort' altitude during the 2012 aerial survey (left) and during the 2014 aerial survey (right) for humpbacks along the GBR. Coastline not shown.

A4.2 Distributions of declination angles

To check if there is substantial differences in the way front and back observers were seeing (i.e., searching) and recording humpback whale sightings, we plotted the distribution of declination angles for both the 2012 and 2014 aerial surveys (Figure A4.2, A4.3).

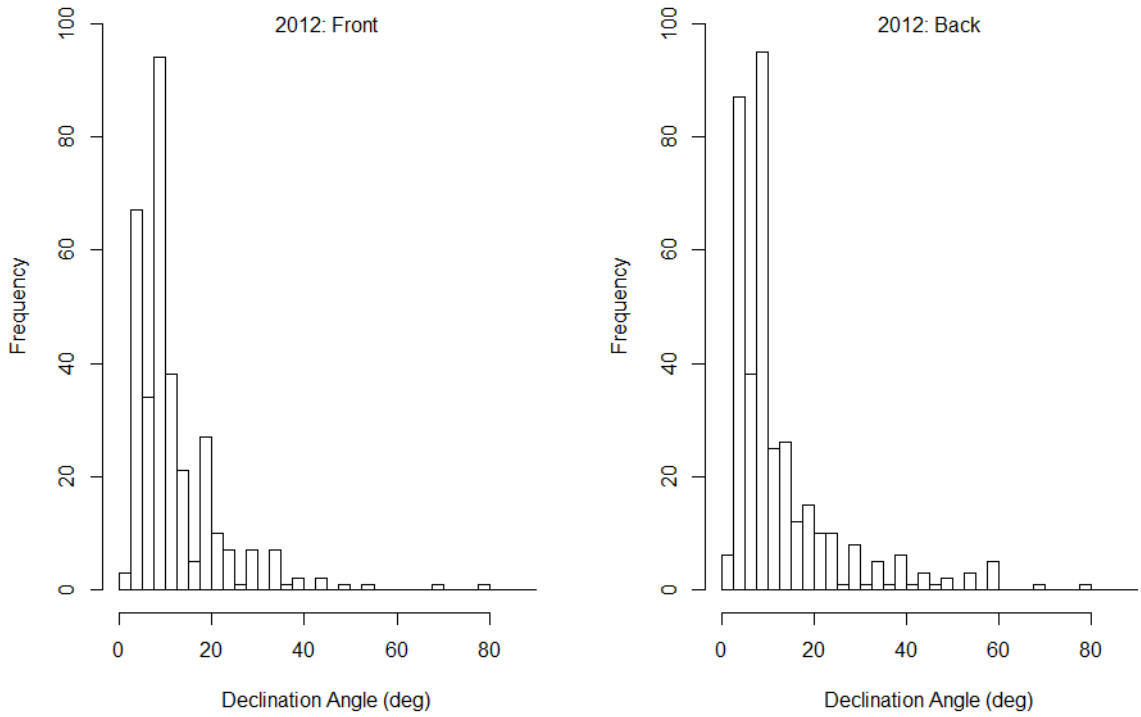


Figure A4.2 Angle of declination (i.e., from the horizontal) for humpback whale sightings for front and back observers during the 2012 aerial survey. Duplicate sightings included in both front and back plots.

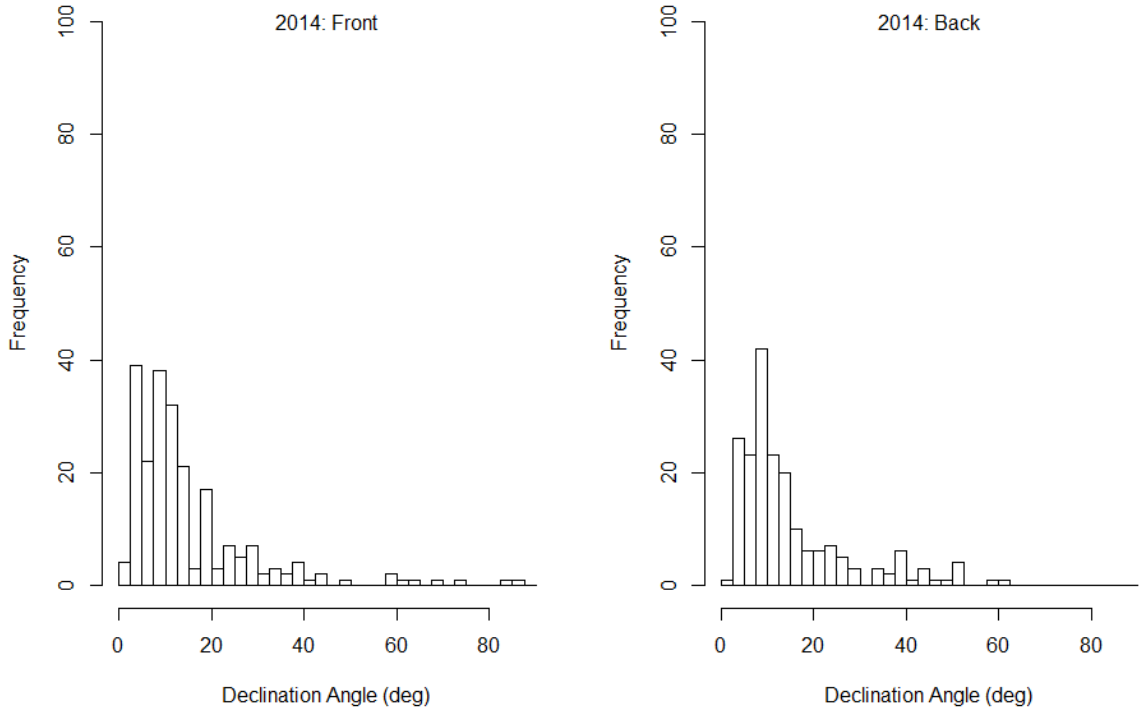


Figure A4.3 Angle of declination (i.e., from the horizontal) for humpback whale sightings for front and back observers during the 2014 aerial survey. Duplicate sightings included in both front and back plots.

A4.3 Distributions of sightings angles

We also investigated the distribution of angle of sightings relative to the heading of the plane (A4.4, A4.5). There was no evidence of any bias in the data.

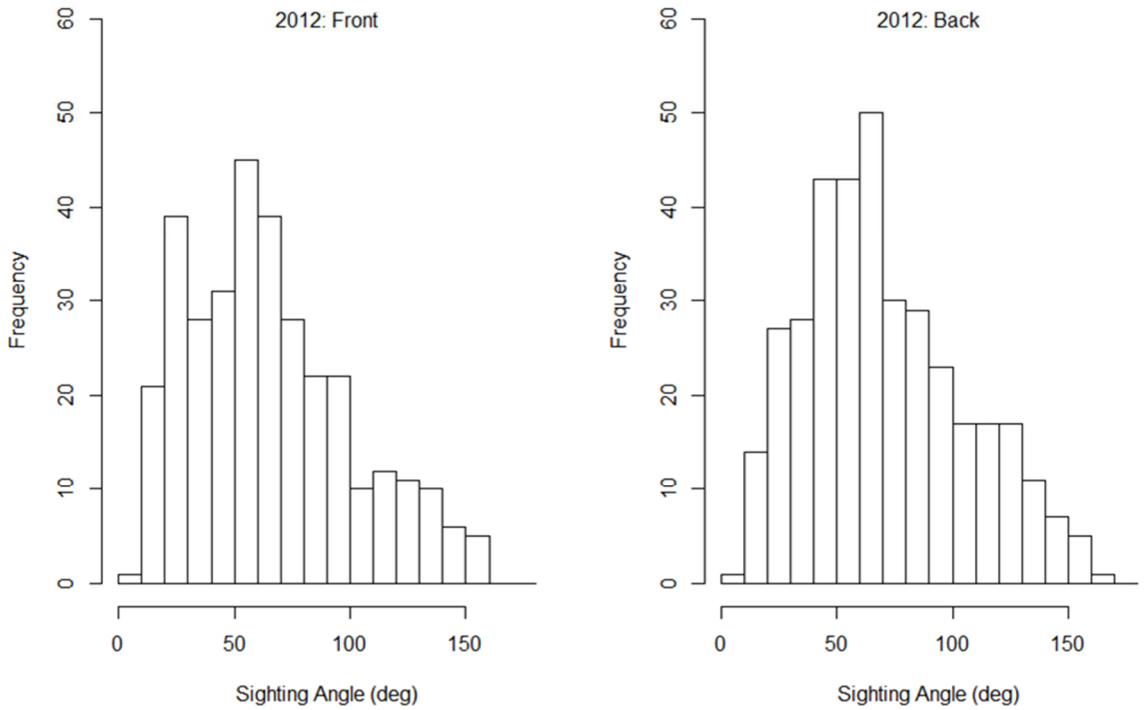


Figure A4.4 Angle of sightings relative to the planes head for humpback whale sightings for front and back observers during the 2012 aerial survey. Duplicate sightings included in both front and back plots.

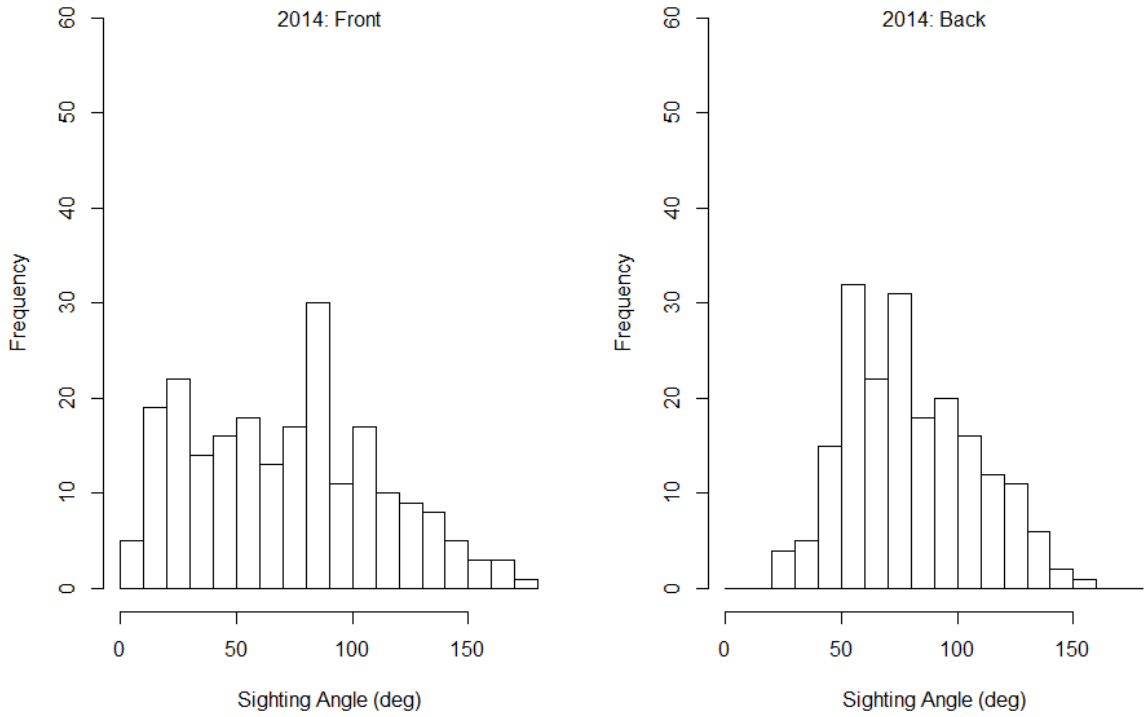


Figure A4.5 Angle of sightings relative to the planes head for humpback whale sightings for front and back observers during the 2014 aerial survey. Duplicate sightings included in both front and back plots.

A4.4 Detection functions

Based on observations in the aircraft, the maximum angle of declination that the rear observers could comfortably observe at is 55-60° from the horizon however the front observer could ostensibly see all the way to the track line owing to the use of bubble windows. These visual limitations need to be considered when setting a left-truncation distance. At an on-effort altitude of around 1000 ft (with some variation during flights of the altitude; see section 0), this corresponds to a strip of about 175 m underneath the aircraft that the back observer can't really see. Considering that, and the actual histogram of perpendicular distances (Figure A4.6, Figure A4.7), a left-truncation distance of 0.2 km was set and a right truncation distance of 4 km. All further analyses used these truncation distances.

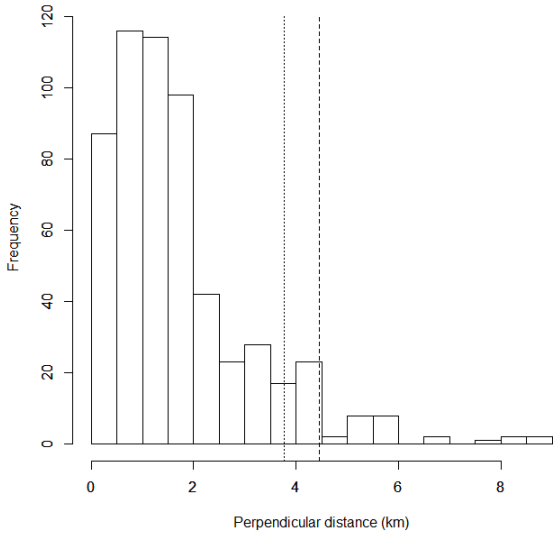


Figure A4.6 Perpendicular distances of humpback whale sightings pooled for front and back observers during the 2012 and 2014 aerial surveys. Duplicate sightings included in both front and back plots.

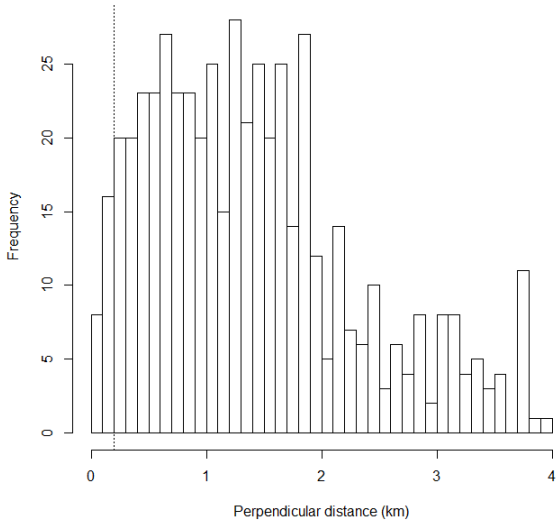


Figure A4.7 Perpendicular distances of humpback whale sightings within 4 km, binned to 100 m. The dotted line is at 200 m, which is the approximate distance at which the rear observers cannot see (based on an estimated upper-declination angle limit of 60°). Data pooled across both survey years.

A4.5 General outputs from sensitivity and other analyses

The following plots are of generalised additive models (GAMs), with single thin-plate spline smoothes of inferred humpback whale densities (estimated via distance analyses) versus the various spatial, physiographic and environmental covariates considered in the whale density modelling for this project. The purpose of these single-variable GAM fits, and resultant plots, were to identify potential function forms and strengths of relationships between whale density and the various covariates. Points around the smoothes are partial residuals.

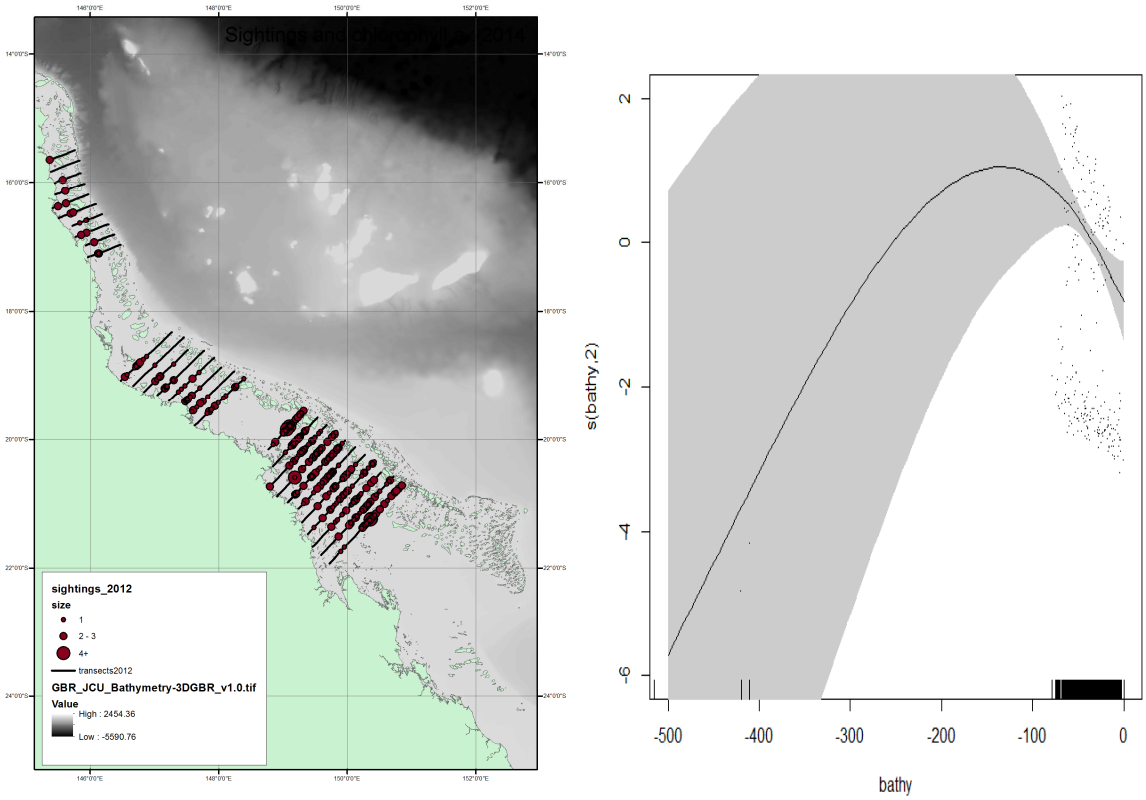


Figure A4.8 Bathymetry and 2012 aerial survey data

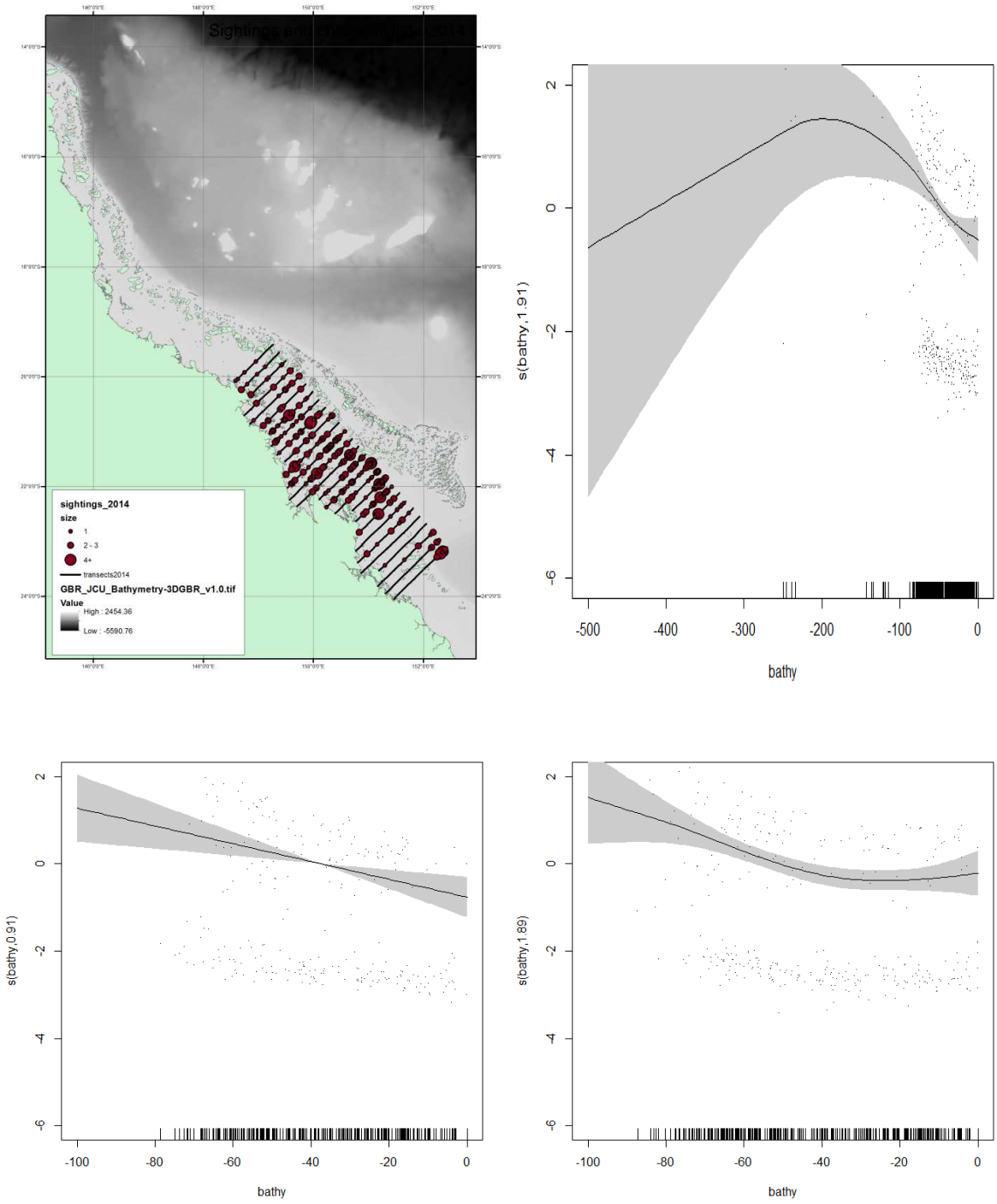


Figure A4.9 Bathymetry and 2014 aerial survey data

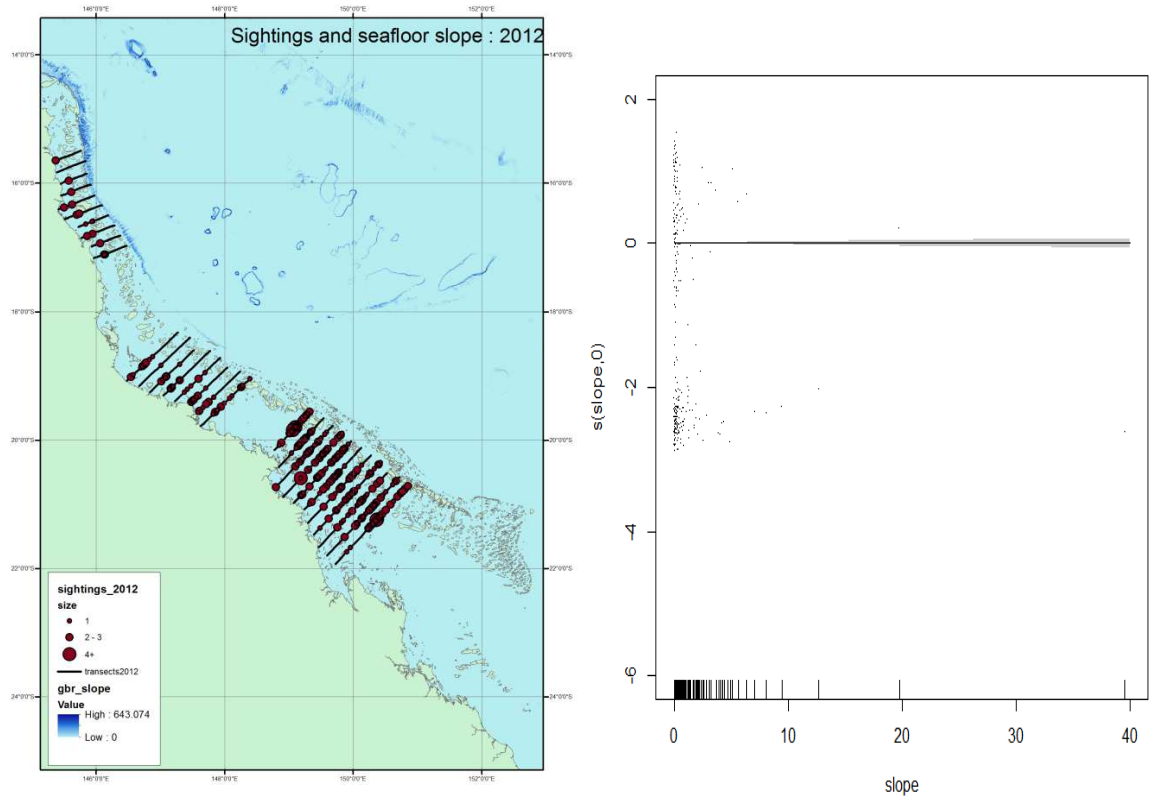


Figure A4.10 Sea floor slope and 2012 aerial survey data

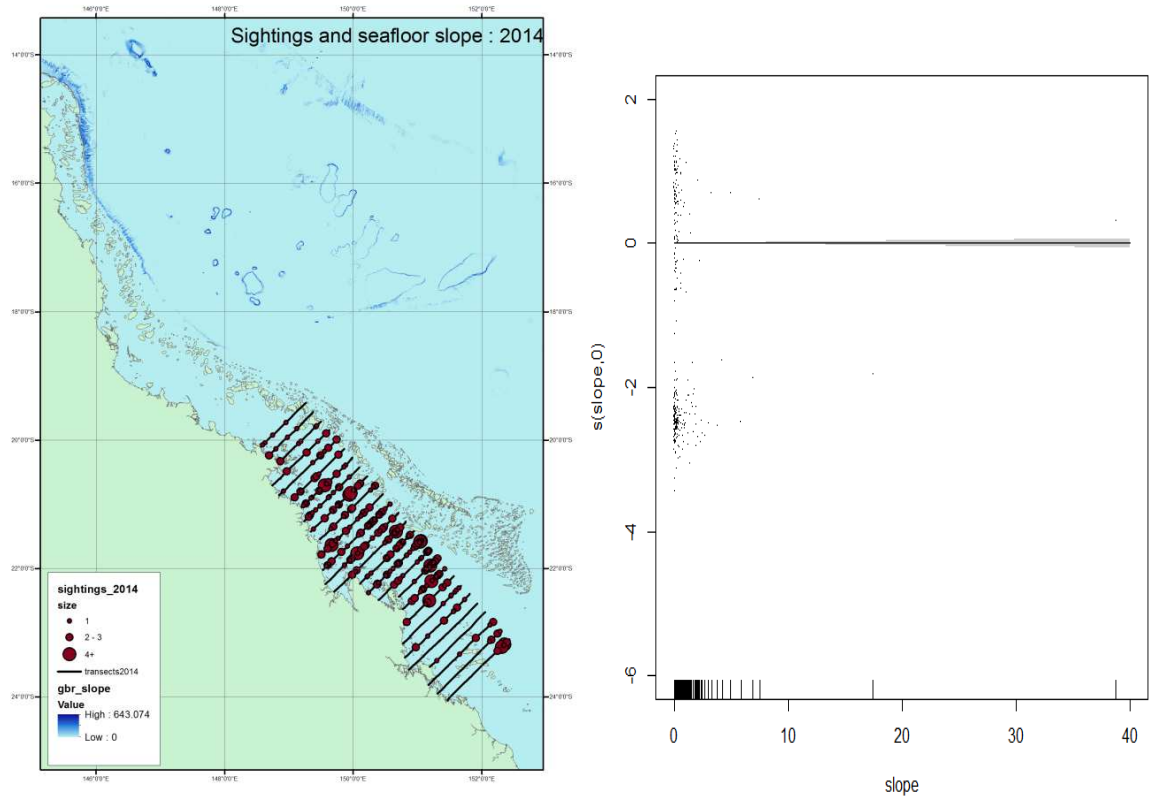


Figure A4.11 Sea floor slope and 2014 aerial survey data

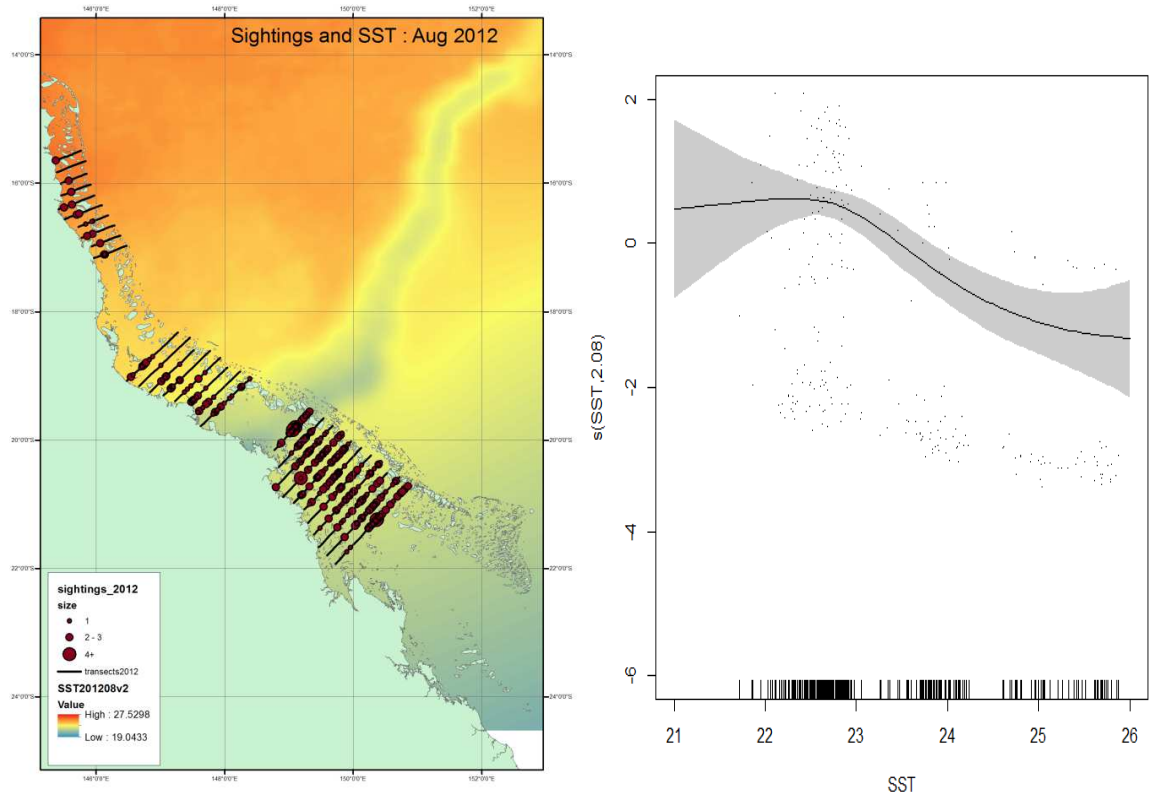


Figure A4.12 SST (°C) and 2012 aerial survey data

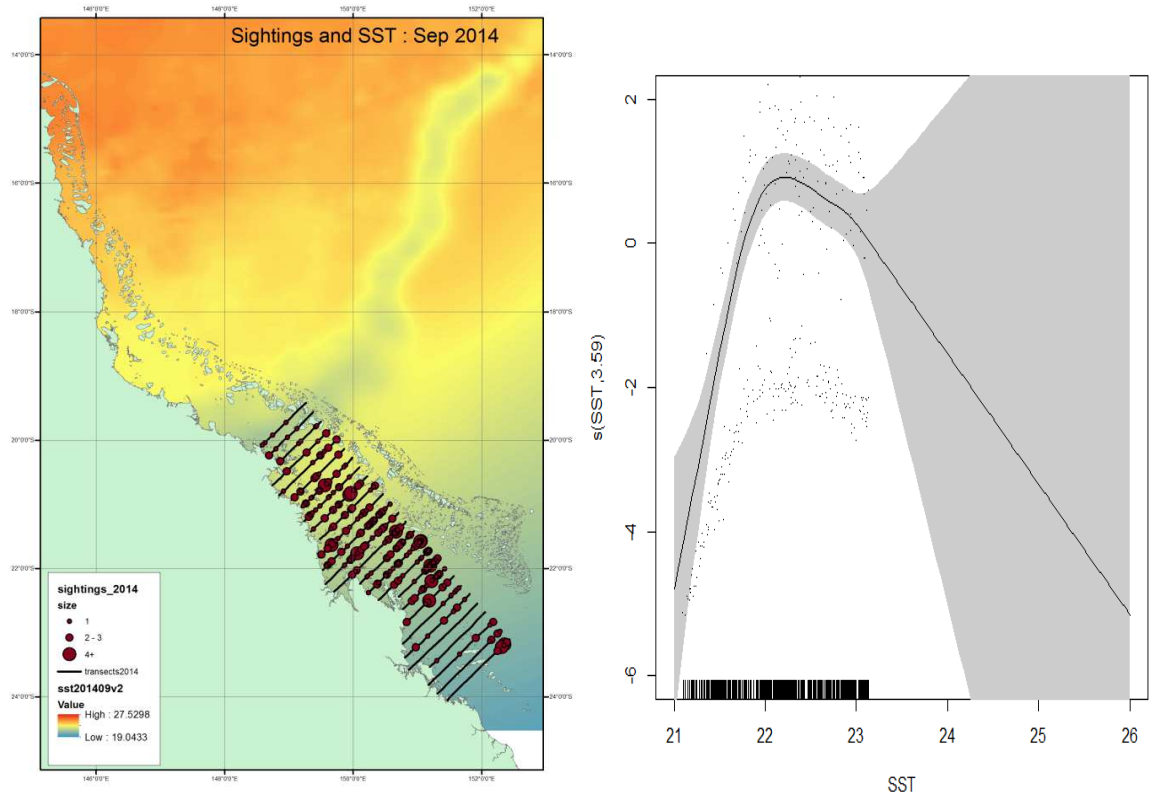


Figure A4.13 SST (°C) and 2014 aerial survey data

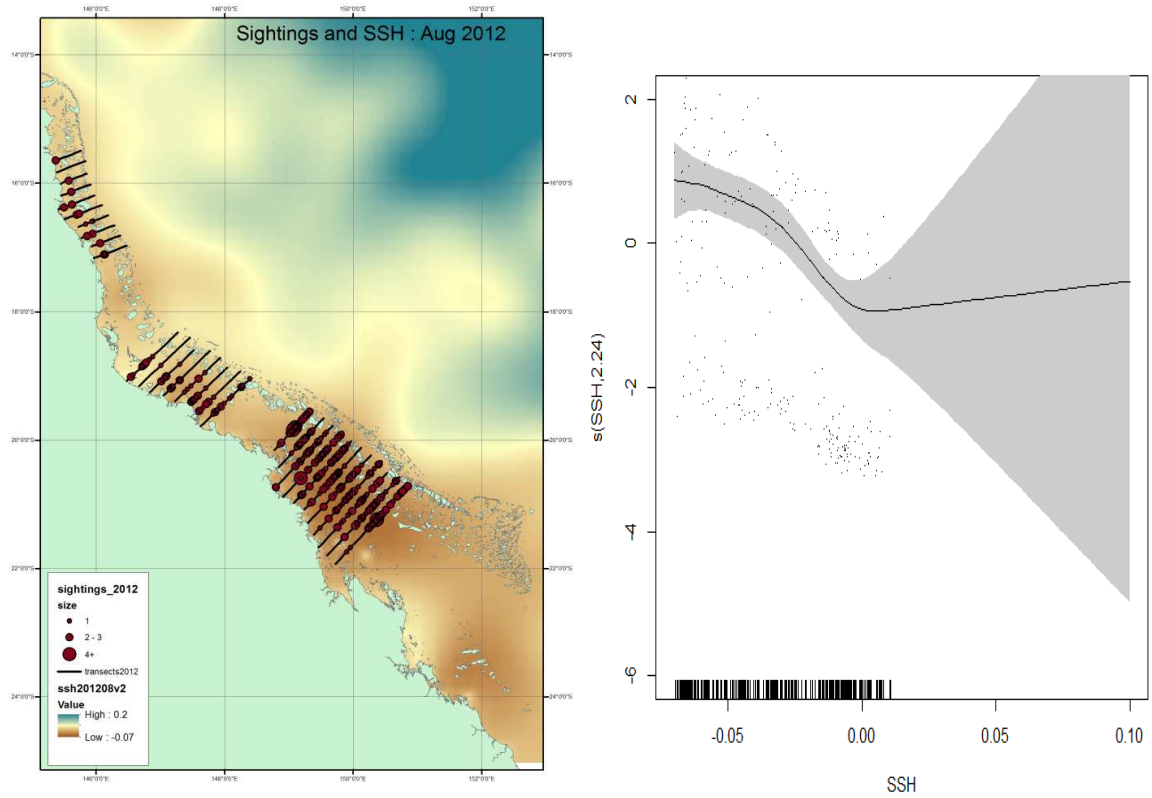


Figure A4.14 SSH anomaly (m) and 2012 aerial survey data

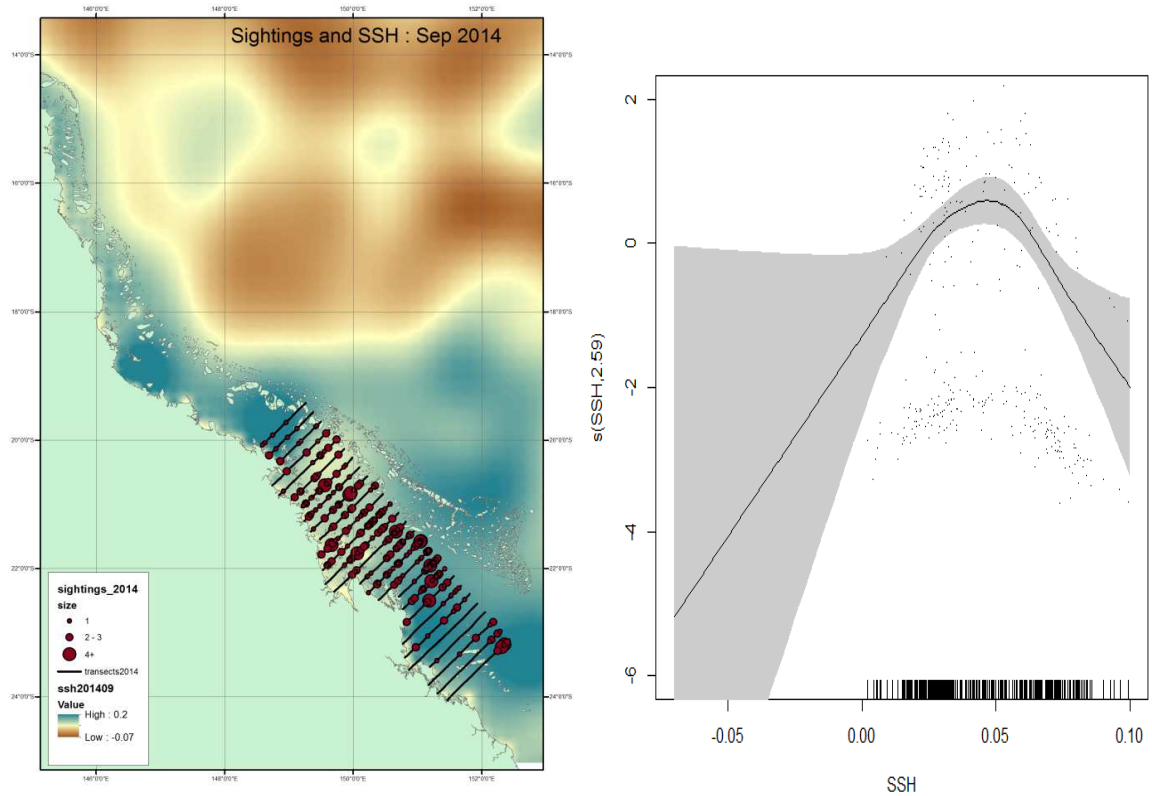


Figure A4.15 SSH anomaly (m) and 2014 aerial survey data

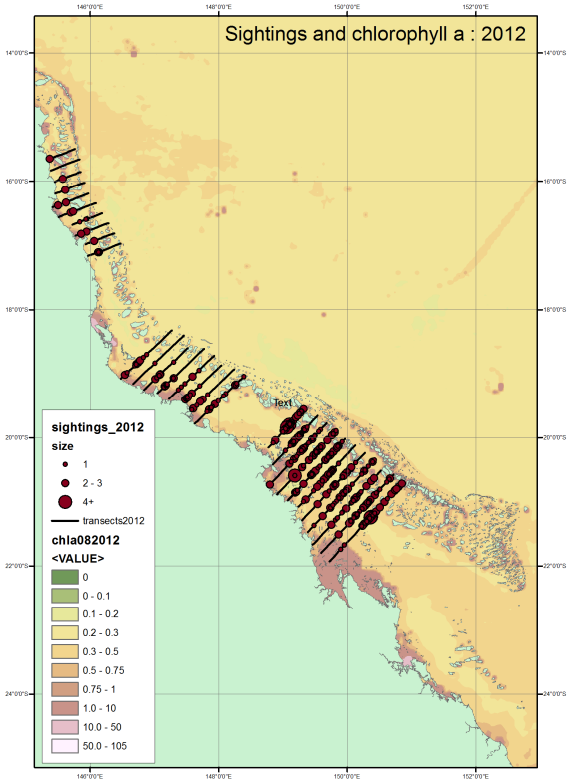


Figure A4.16 Chlorophyll *a* and 2012 aerial survey data

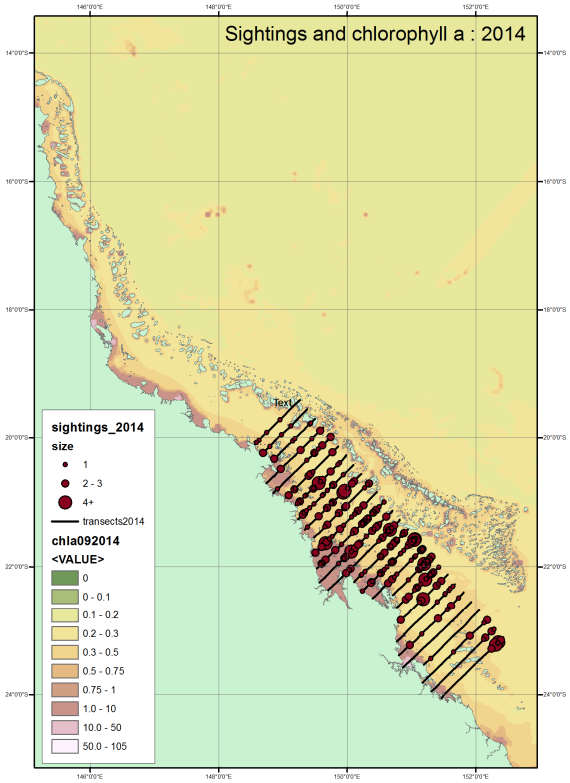


Figure A4.16 Chlorophyll *a* and 2014 aerial survey data

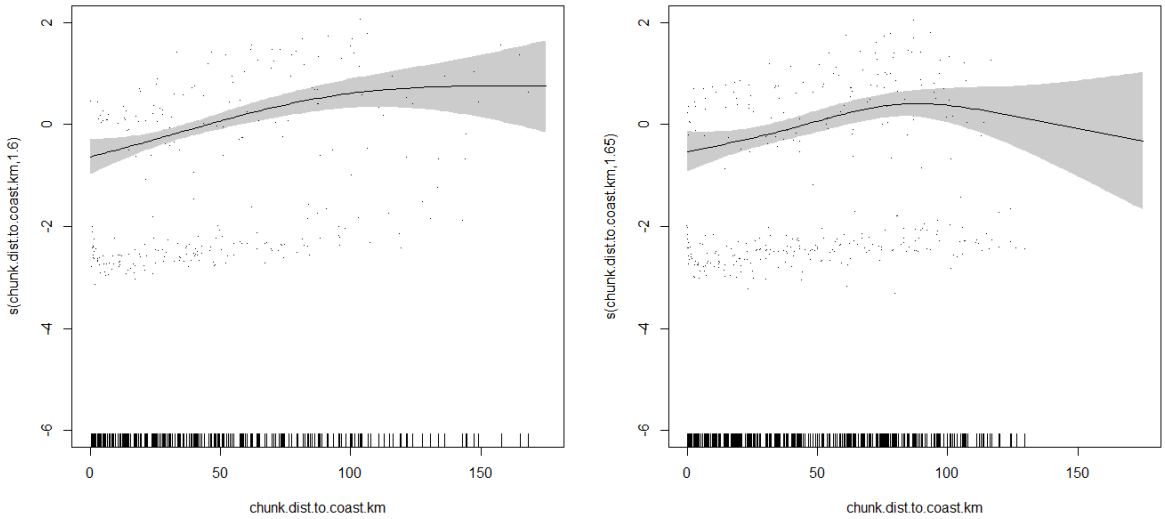


Figure A4.18 10 km line segments and distance to the coast for the 2012 survey (left panel) and the 2014 survey (right panel).

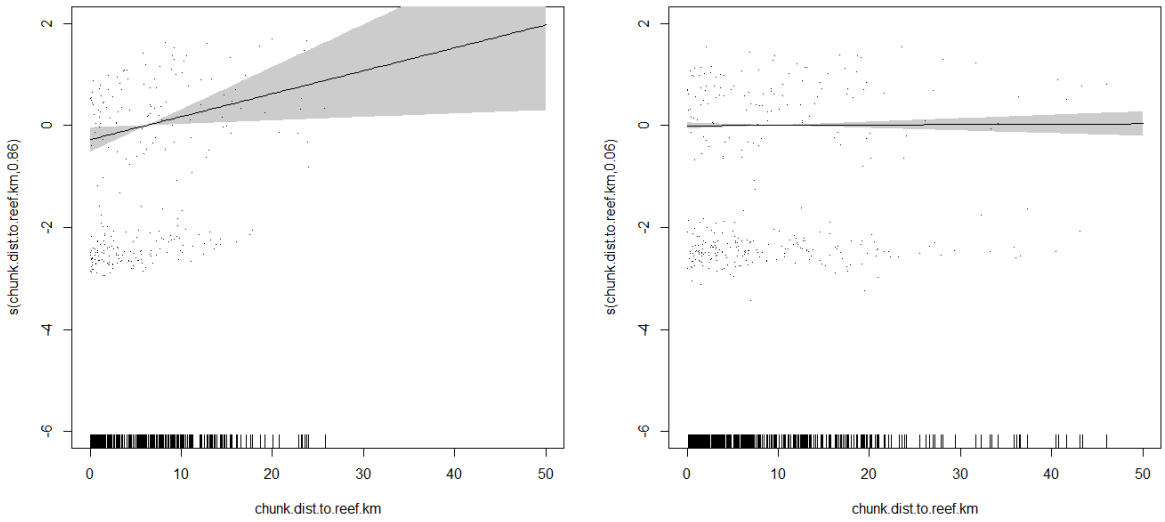


Figure A4.19 10 km line segments and distance to the nearest reef for the 2012 survey (left panel) and the 2014 survey (right panel).

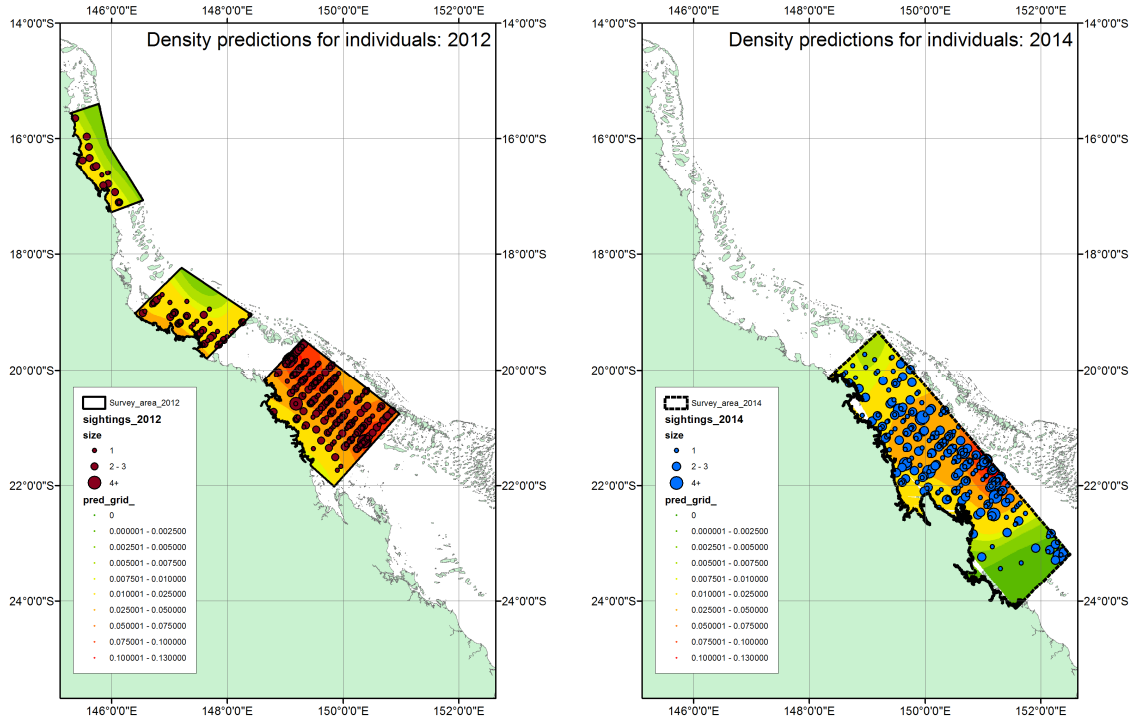


Figure A4.20 Density predictions (individuals per 1 km^2) for all humpback whales from 2012 and 2014 aerial surveys

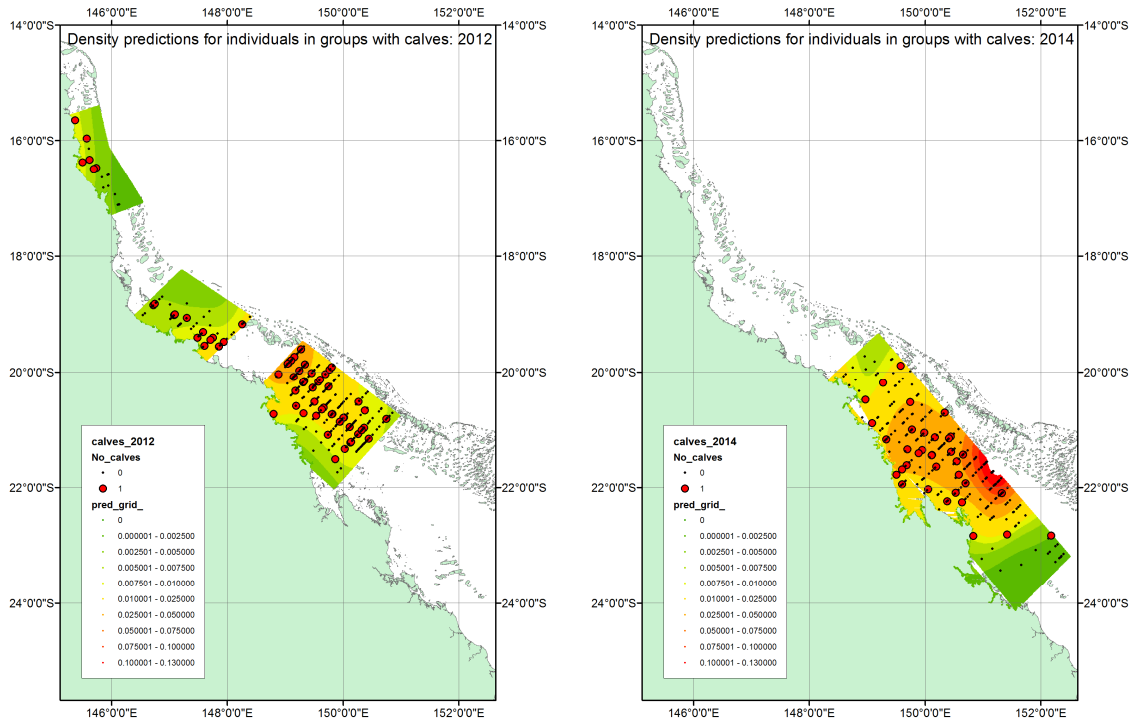


Figure A4.21 Density predictions (individuals per 1 km^2) for individuals in groups with calves from 2012 and 2014 aerial surveys

Appendix 5 Additional details relating to the estimation of the relative probability of a fatality

5.1: Aerial survey and vessel availability

In a typical aerial survey analysis the formula provided by Barlow et al. (1988) is used,

$$\Pr(\text{Visible}) = \frac{s+t}{s+d}$$

Where *s* is average time near the surface, *d* is the average time at depth and *t* is the time the whale is within the field of view of aerial observers (e.g., taking into account aircraft speed and altitude).

Hence the absolute abundance *A*, is given by

$$\text{A} = \frac{A_{observed}}{\Pr(\text{Visible})}$$

So in our case if we were to adjust for the aerial survey availability bias and then multiply by $\Pr(\text{w}_{Depth} > \text{v}_{Draft})$ we would have.

$$\frac{A_{observed}}{\Pr(\text{Visible})} \times \Pr(\text{w}_{Depth} > \text{v}_{Draft})$$

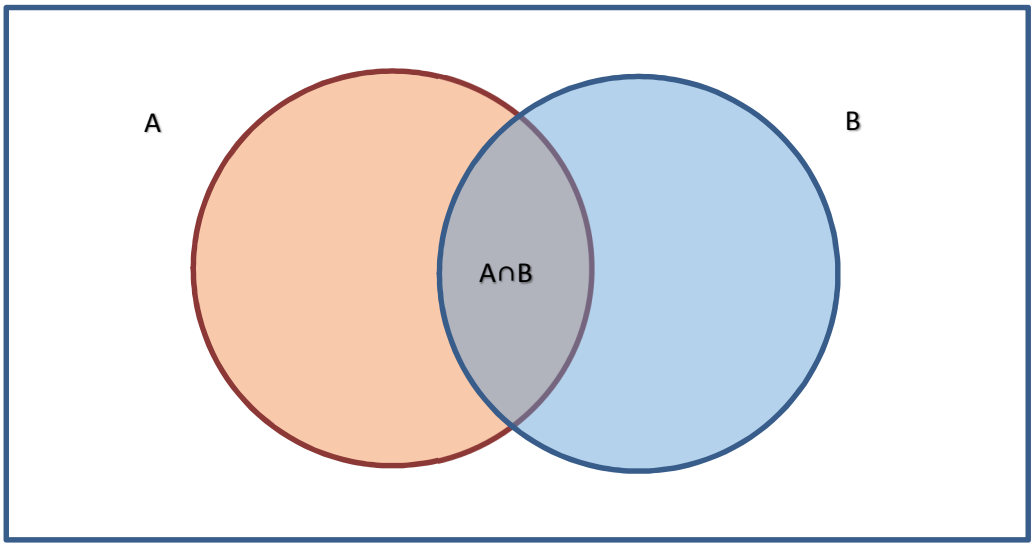
$$= A_{observed} \frac{s+d}{s+t} \times \frac{s}{s+d}$$

$$= A_{observed} \frac{s}{s+t}$$

There is not much information in the literature for surface duration of humpback whales. Dolphin (1987) found for Alaskan humpback whales in the feeding grounds typical non-feeding surface duration was 2.6 minutes and since aerial survey availability is typically of the order 7-14 seconds, you can see the adjustment will be approximately 90-95%.

5.2: Probability Rule

If we are considering two independent events A and B, each with probabilities given by Pr(A) and Pr(B), the Pr(A and B) is the intersection on our Venn diagram $\Pr(A \cap B)$.



The probability of (A or B) is given by

$$Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$

Since if we do not take away the $\Pr(A \cap B)$ we will double count it. $\Pr(A \cap B)$ since A and B are independent $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$. So

$$Pr(A \; or \; B) = \Pr(A) + \Pr(B) - \Pr(A)\Pr(B)$$

For more than two events this generalises to,

$$\Pr(A_1 \; or \; A_2 \; or \; \dots \; or \; A_n) = 1 - \prod_{i=1}^{n}[1 - Pr(A_i)]$$

5.3: Probability of a fatality given w whales in a cell

If we wish instead of looking at the expected number of fatalities we can look at the probability of fatality, which may be useful. So that gives the probability proportional to the probability of a fatal strike for a single whale in the grid cell, to estimate for any/all whales in the grid cell we use the same 'or' probability rule again

$$\Pr(Fatality) = 1 - \prod_{w=1}^{W}[1 - Pr(Fatality_w)] \quad \text{for } w > 0, 0 \text{ otherwise}$$
$$= 1 - [1 - Pr(Fatality_w)]^W$$

Now this causes an issue as although in reality *W* corresponds to the number of animals in the grid cell, and hence is discrete whole numbers, in practice the whale model provides us with a continuous numbers. The repercussion of this is that the concept of OR breaks down (e.g., if the abundance for a cell is 1.5, then how does this correspond to an OR of events). To conceptually work around this difficultly we propose to consider partial whales as simply whales that were not in the cell for the whole season (e.g., 0.5 would equate to a whale (not necessarily the same whale) in the cell for only half the season[12]). So, taking this interpretation, partial whales will just have the effect of including a multiplier to the $Pr(Fatality_w)$ in one of the OR events. To demonstrate why this is so, consider our 0.5 example, since the animal is only available to be struck for 50% of the available time it would be expected to be half as likely to be struck.

Now we have calculated the $Pr(Fatality_w)$ for each cell, as per our equations, the key point is we defined this as the probability given a single whale was present in the grid cell of one of the vessels present having a collision with the animal. So now we have to adjust this for the probability that the whale is actually in the cell so

$$Pr(Fatality_w) \times \Pr(whale \; in \; the \; cell)$$

which corresponds to the fractional abundance, so in this example would be 0.25.

To handle partial whale number >1. If we take the probability equation for the whole number of whales is given as before by

$$\Pr\left(Fatality_{1,..\lfloor W \rfloor}\right) \propto [1 - Pr(Fatality_w)]^{\lfloor W \rfloor} \quad \text{for } w > 0, 0 \text{ otherwise}$$

where $\lfloor W \rfloor$ denotes the floor of the number W. Then the probability of the partial remainder is given by

$$\Pr\left(Fatality_{W-\lfloor W \rfloor}\right) \propto (W - \lfloor W \rfloor) \times Pr(Fatality_w)$$

So the total probability is simply an OR between these two probabilities

[12] This touches on another point that we do not assume a whole whale value corresponds to a whale being in a cell for all of the season, but rather that whales come and go but there is on average of that many whales in the grid cell

$$\Pr(Fatality) \propto \Pr\left(Fatality_{1,..\lfloor W \rfloor}\right) + \Pr\left(Fatality_{W-\lfloor W \rfloor}\right) - \Pr\left(Fatality_{1,..\lfloor W \rfloor}\right) \times \Pr\left(Fatality_{W-\lfloor W \rfloor}\right)$$

Appendix 6 AMSA Craft Tracking System (CTS) metadata

Table A6.1 AMSA Craft Tracking System AIS Metadata

CSV Table Field name	Type	Description
OBJECTID	Object ID	ArcGIS primary key
POSITION_ID	Double	CTS.POSITION table primary key
CRAFT_ID	Double	CTS unique identifier for each vessel
CRAFT_REPORTING_AGENT_ID	Double	Data source code: 20 = AIS
CRAFT_TYPE_ID	Double	Craft type code: 1 = *vessel*
FIX_ID	Double	Join field to other CTS tables
LON	Double	Longitude in decimal degrees
LAT	Double	Latitude in decimal degrees
POSITION_TIME	Date	UTC timestamp of vessel position report
CREATED_TIME	Date	UTC timestamp when vessel position report was written to CTS database
ALTITUDE_METRES	Double	Aircraft altitude. Not populated
COURSE_DEGREES	Double	Course over ground in decimal degrees
HEADING_DEGREES	Double	Heading in decimal degrees. Not available
SPEED_KNOTS	Double	Speed over ground in knots
CRAFT_TYPE	Text	Vessel type
CRAFT_SUBTYPE	Text	Vessel sub-type
LENGTH_METRES	Double	Vessel length in metres
BEAM_METRES	Double	Vessel beam in metres
MID_CODE	Long Integer	MMSI country identification digit
ORIGIN	Text	Craft origin. Not available for AIS records
DESTINATION	Text	Vessel destination
CLS	Text	AIS Class A or B
MMSI	Long Integer	Maritime Mobile Service Identity (MMSI) number
IMO	Long Integer	International Maritime Organization (IMO) number
REGISTRATION	Text	Vessel call sign
NAME	Text	Vessel name
SOURCE_DETAIL	Text	AIS satellite identifier
POOL	Text	For aircraft use only
DRAUGHT_METRES	Double	Vessel draught in metres. Please note, vessel draught data is only available for vessel position records from 21 June 2013 onwards

Appendix 7 Alternate co-occurrence results

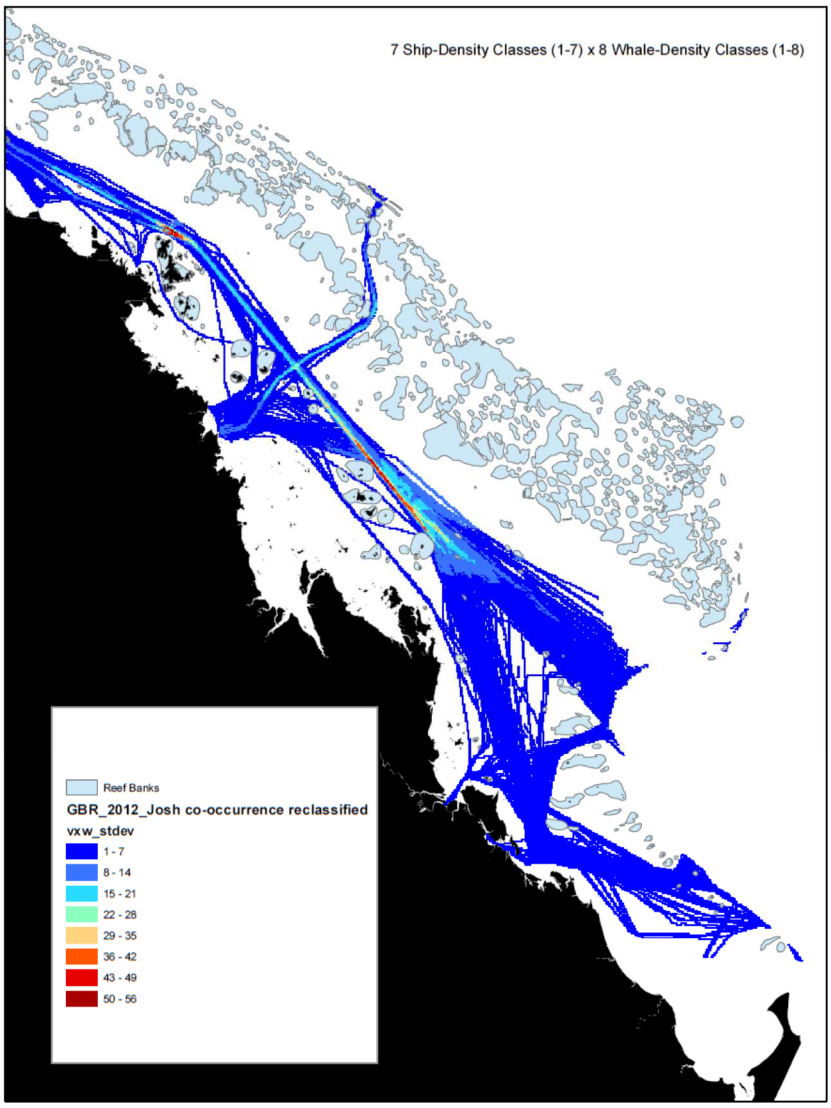


Figure A6.1 Co-occurrence using data categories based on standard-deviation.

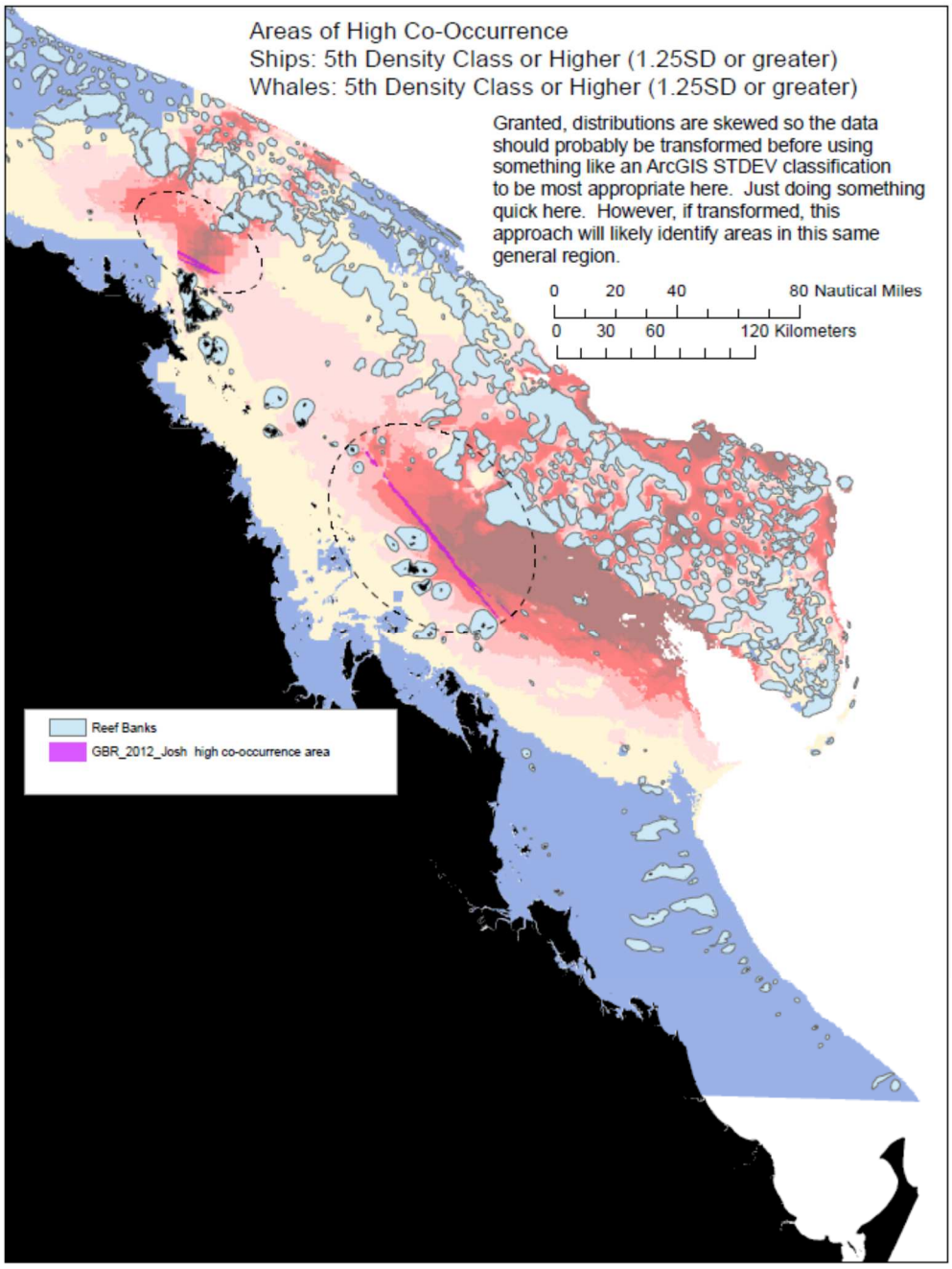


Figure A6.2 High co-occurrence based on highlighting areas where the highest categories in whale and shipping occur.