western australian
marine science institution

Fine Scale Spatial Models: *Spatial predictive models of mixed benthic assemblages across the offshore Kimberley region*

Marji Puotinen[1,2], Ben Radford[1,2], Karen Miller [1,2]

[1]Australian Institute of Marine Science, Perth, Western Australia

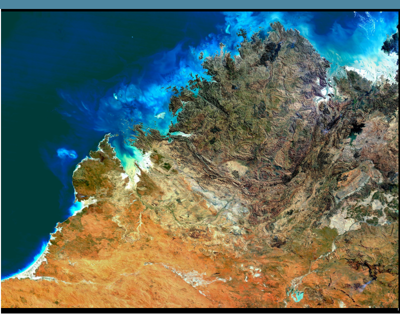
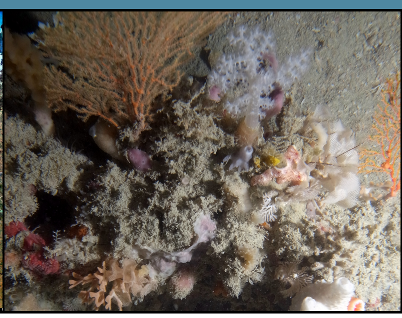
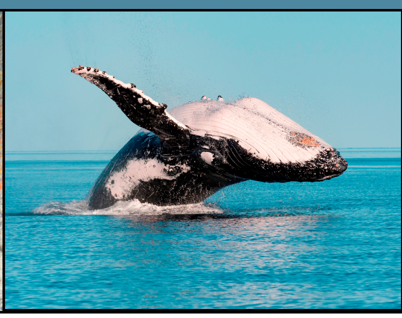
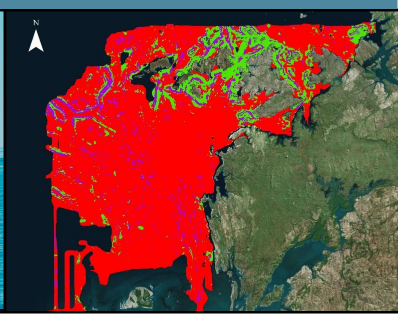
[2]Western Australian Marine Science Institution, Perth, Western Australia

WAMSI Kimberley Marine Research Program

Final Report

Subproject 1.1.1.3

September 2018

WESTERN AUSTRALIAN MARINE SCIENCE INSTITUTION

Australian Government

Australian Institute of Marine Science

WAMSI Kimberley Marine Research Program

Initiated with the support of the State Government as part of the Kimberley Science and Conservation Strategy, the Kimberley Marine Research Program is co-invested by the WAMSI partners to provide regional understanding and baseline knowledge about the Kimberley marine environment. The program has been created in response to the extraordinary, unspoilt wilderness value of the Kimberley and increasing pressure for development in this region. The purpose is to provide science based information to support decision making in relation to the Kimberley marine park network, other conservation activities and future development proposals.

Front cover images (L-R)

Image 1: Satellite image of the Kimberley coastline (Image: Landgate)

Image 2: Mixed brozoans soft corals sponges and seafans on a rock (Image: AIMS)

Image 3: Humpback whale breaching (Image: Pam Osborn)

Image 4: Figure 18: Mixed benthic habitat classification map of the Southern study area (C1).

Year of publication: September 2018

Metadata: https://apps.aims.gov.au/metadata/view/49791f9e-c8a9-4844-ae31-165f9761cd0f

Citation: Puotinen M, Radford B, Miller K (2018). Fine scale spatial models: Spatial predictive models of mixed benthic assemblages across the offshore Kimberley region report subproject 1.1.1.3 prepared for the Kimberley Marine Research Program, Western Australian Marine Science Institution, Perth, Western Australia, 42pp.

Author Contributions: All authors contributed to the drafting of this text.

Corresponding author and Institution: Enquires specific to individual chapters should be addressed to the relevant institutional lead investigators. General enquiries should be addressed to Dr Andrew Heyward, Australian Institute of Marine Science, a.heyward@aims.gov.au.

Funding Sources: This project was funded (commissioned) by the Western Australian Marine Science Institution as part of the WAMSI Kimberley Marine Research Program, a $30M program with seed funding of $12M provided by State government as part of the Kimberley Science and Conservation Strategy. The Program has been made possible through co-investment from the WAMSI Joint Venture partners and further enabled by data and information provided by Woodside Energy Ltd.

Competing Interests: The commercial investors and data providers had no role in the data analysis, data interpretation, the decision to publish or in the preparation of the manuscript. The authors have declared that no competing interests exists.

Kimberley Traditional Owner agreement: This research was enabled by the Traditional Owners through their advice, participation and consent to access their traditional lands.

Acknowledgements: The WAMSI Benthic Biodiversity project 1.1.1 undertook surveys in Western Australian waters along the Kimberley coast and also further offshore in Commonwealth waters. We are grateful to Traditional Owners (the Wunambal Gaambera and Dambimangari people) of the Kimberley sea country where some of the studies were done. We thank the Wunambal Gaambera Aboriginal Corporation, the Uunguu Rangers and the Dambimangari Aboriginal Corporation for their assistance, advice and participation in the Benthic Biodiversity project. The authors wish to express our thanks to the Masters and crews of RV Solander and RV Linnaeus for the support and many courtesies extended to the research teams during the field surveys.

Collection permits/ethics approval: This study was part of WAMSI Benthic Biodiversity project 1.1.1. Sampling of representative biota for biodiversity characterisation and lodgement of taxonomic reference material was undertaken under permits; WA Fisheries Exemption No. 2547, 2677, 2721; WA Department of Parks and Wildlife CE 004795, SF010234, SF10627, SF010720; Commonwealth Department of Environment and Energy AU-COM2016-326;

.

Contents

Executive Summary

Benthic habitat information was used in combination with multibeam data to develop spatial predictive models of mixed benthic assemblages, showing the probable distribution of various mixed benthic habitat classes.

Greatly improved bathymetry data for each of the three survey regions were used to generate maps showing the probable distribution of various mixed benthic habitat classes. These models also provide a means to estimate the percentage area covered by each class of habitat in the survey region.

In general, validation of the benthic habitat models showed they performed well. Areas of high model accuracy were more widespread across the Southern than the Central region. The overall classification accuracy for the Northern study area was notably lower than for either Southern or Central region, and areas of low accuracy were much more widespread spatially than areas of high accuracy.

The detailed spatial information that mapping classification accuracy provides can be used for making informed spatial allocation decisions for conservation such as determining the placement of protective zones. Ideally, it would be preferable to make use of these local area models to assist in management decisions, such as a spatial zoning and use approach, for areas that were predicted with high rather than low confidence. The project's maps of classification accuracy enable such decision processes for the first time.

1 Introduction

The spatial distribution of key benthic communities across the remote offshore Kimberley region of north-west Australia is still poorly understood. Logistical constraints have limited direct field data collection to only a small portion of the area contained within existing and proposed management areas, as has been generally true for marine organisms around the globe (Edgar et al. 2016). To address this huge data gap, we aimed to predict the spatial distribution of key classes of bottom-dwelling organisms across the region in as much spatial detail as robustly possible while making sure that the resulting maps were:

- Ecologically meaningful on relevant spatial and temporal scales,
- Sufficiently accurate for the intended use, and
- Communicated clearly in terms of their likely errors and uncertainties.

Well established ecological theory (Elith et al. 2008, Elith et al. 2009, Holmes et al. 2006, Leathwick et al. 2008, Pittman et al. 2009) details how seafloor physical properties act as both direct and indirect drivers of landscape scale ecological processes on the benthos. Implementing this knowledge within a GIS via robust statistical modelling techniques, we built habitat maps predicting where ecologically significant habitats such as coral, macro-algae, sponges and other invertebrates are likely to be found. Such maps will be of significant value for environmental planning and management.

Our approach focuses on identifying and mapping the potential drivers of the spatial distribution of habitats (or related surrogates), and using field observations of the spatial distribution of those biota to predict where biota are likely located in areas not surveyed (Brown et al. 2011, Holmes et al. 2008). This approach is particularly useful for large, remote areas where practicality limits the number of samples that can be collected, such as NW Australia's marine estate (for example, see Huang et al. 2014). Further, we explore and optimise the spatial scale at which we represent potential drivers in our models, which is essential to provide ecologically interpretable and robust models because spatial scale can make a big difference to the results (for example, Monk et al. 2011).

Careful survey planning was undertaken to ensure that the data collected was sufficient to build spatial predictive models over as large an extent of the region as possible while still yielding robust results (see 1.1.1.1 Chapters 2 and 3).

2 Methods and Results

We collected the hydro-acoustic and biological data needed for spatial predictive modelling of benthic assemblages over three field campaigns (see Figure 1; Figure 4, 1.1.1.1 Chapter 1; Figure 4, Synthesis):

- Southern (C1): within part of Camden Sound,
- Central (C2): between Bigge Island and the Maret Islands, and
- Northern (C3): between East Holothuria Reef, the Eclipse Islands and the Bougainville Peninsula.

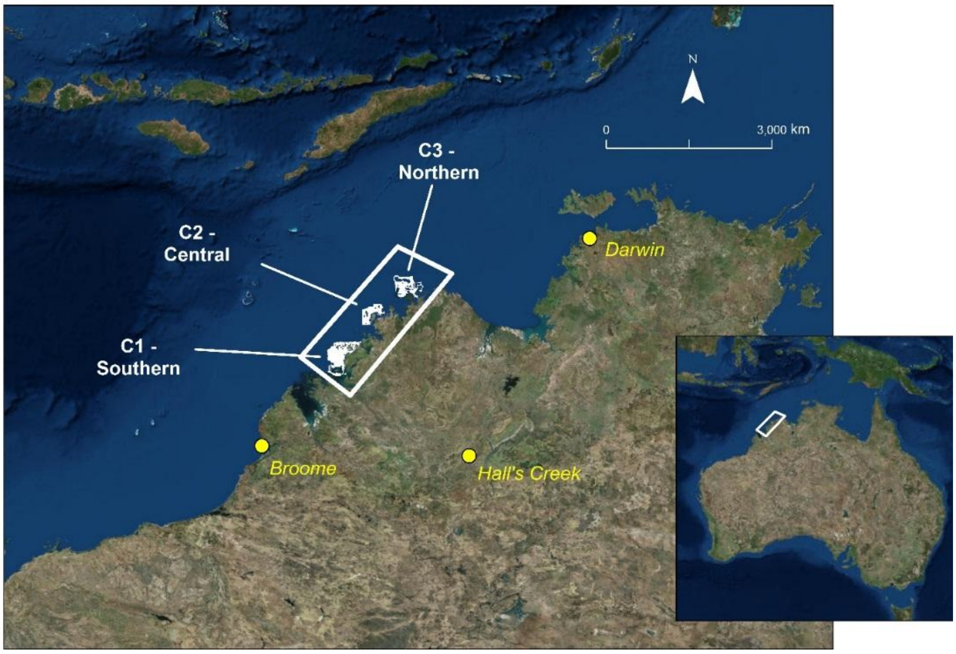


Figure 2: WAMSI study areas (shaded white) in the north-west Australian Kimberley region surveyed during campaigns c1, c2 and c3.

The largest of these was Southern (4,166 km^2), which covered roughly twice as much area as Central (1,824 km^2) or Northern (2,513 km^2).

We built spatial predictive models for each of the three sections of the offshore Kimberley using the following process (Figure 2).

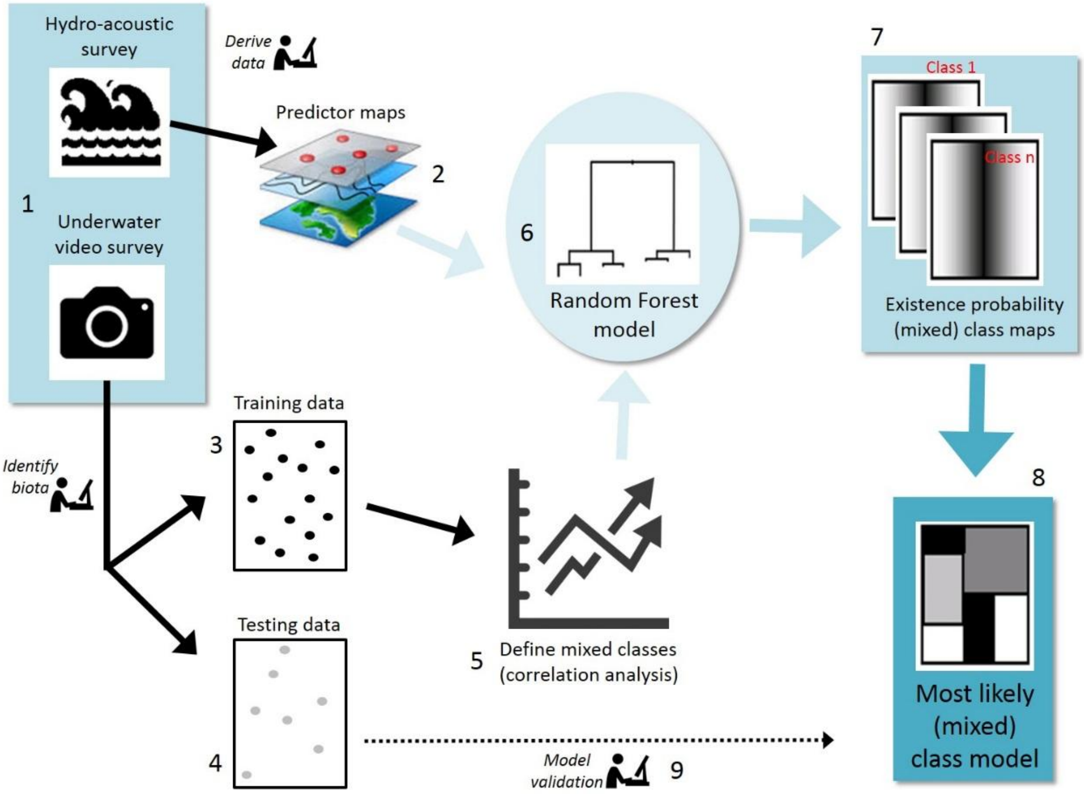


Figure 3: Work flow for building spatial predictive models of benthic communities from field data collected via three field campaigns in the offshore Kimberley region.

2.1. Developing predictors

Developing environmental surrogates to attempt to predict the existence and abundance of classes of benthic organisms (step 2, Figure 2) is possible with high resolution bathymetric data (Brown et al 2011). Where such data do not exist in an area of interest, they can be developed from multi-beam sonar data via hydro-acoustic surveys (Holmes et al 2008, Lehmann et al. 2002- step 1 on Figure 2). For this project, we collected multi-beam data, sourced existing data, and used spatial interpolation to create a high resolution (25 m) bathymetry dataset for each study area (Figures 3-5). The study areas varied in the range of depths they covered, with Southern (Figure 3) covering the greatest range (0.5 - 101 m), followed by Northern (8 - 91 m; Figure 4) and Central (4.5 - 65 m; Figure 5). The mean depth was similar for the Central ($41^{+}8$ m) and Northern ($43^{+}13$ m) study areas, while it was slightly shallower for Southern ($33^{+}11$ m).

In these depth ranges it would be reasonable to expect a variety of predominantly filter feeding and detritivorous species, with benthic primary producers restricted to the shallowest areas (see 1.1.1.2 & 1.1.1.4).

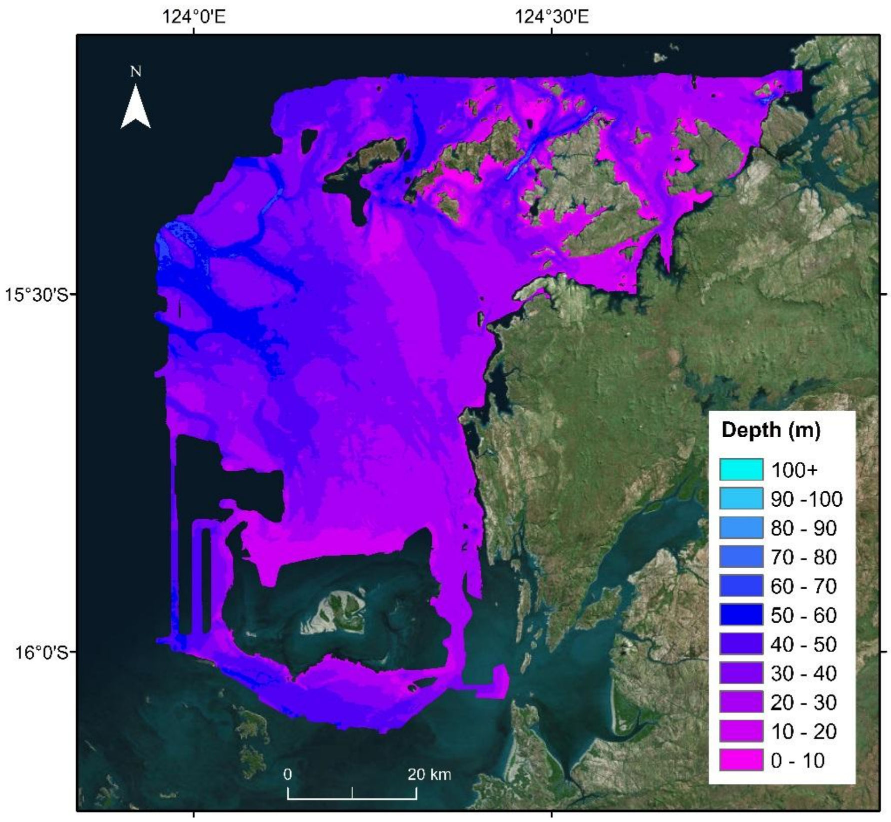


Figure 4: Bathymetry for Southern at 25m spatial resolution, interpolated from multi-beam sonar data collected as part of this project.

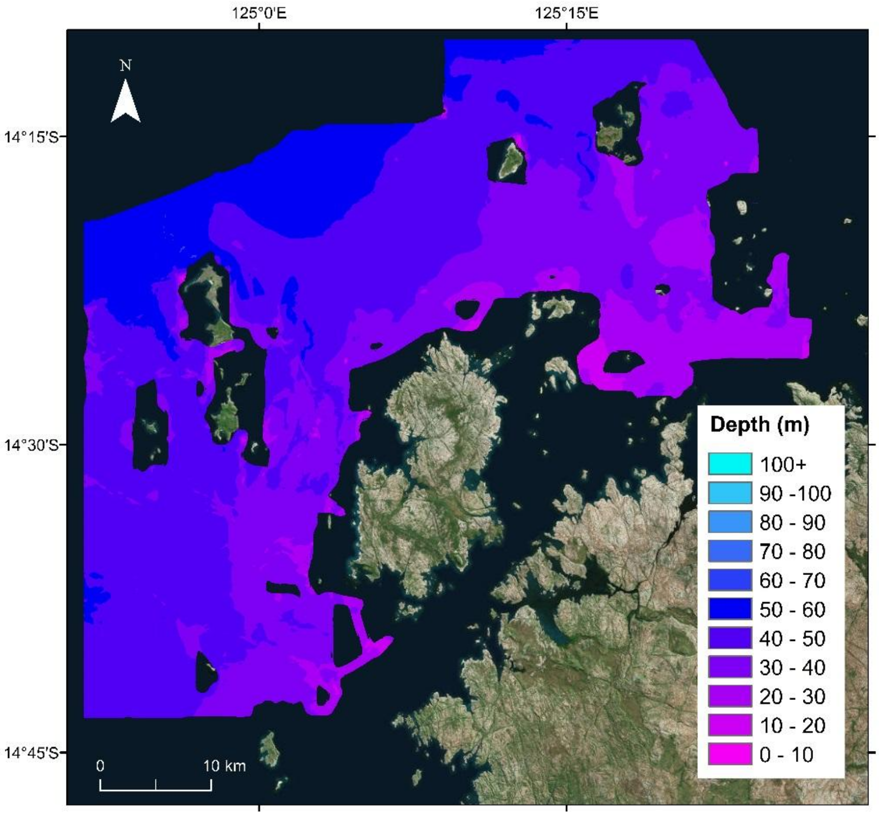


Figure 5: Bathymetry for Central at 25 m spatial resolution, interpolated from multibeam data collected as part of this project.

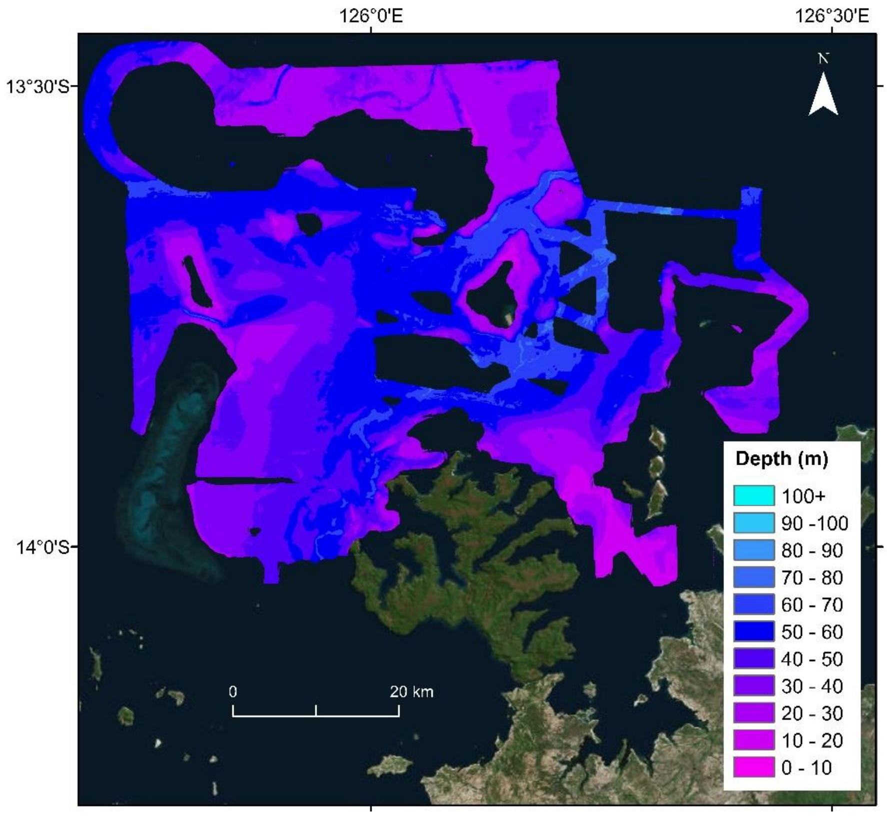


Figure 6: Bathymetry for Northern at 25 m spatial resolution, interpolated from multibeam data collected as part of this project as well as existing multibeam data.

From the bathymetric data, we developed the following potential predictor variables of benthic habitat (Table 1; Figure 2, step 2):

Table 1: Predictor variables used in benthic habitat models.

Benthic habitat predictor variable	Description	Predictor variable prefix(es)
Bathymetry	Water depth in metres, interpolated from multi-beam data to a 25 m resolution	c_depth
Aspect	Azimuthal direction of the steepest slope, calculated for a 3 x 3 pixel neighbourhood	c_aspect
Overall curvature	Combined index of profile and plan curvature (see below)	c_curv
Profile curvature	Second derivative of elevation: concavity/convexity parallel to the slope, calculated for a 3 x 3 pixel neighbourhood	c_prof
Plan curvature	Second derivative of elevation: concavity/convexity perpendicular to the slope, calculated for a 3 x 3 pixel neighbourhood	c_plan
Depth range across various spatial neighbourhoods	Maximum minus minimum depth within spatial neighbourhoods equivalent in width to: 5m, 10m, 20m, 25m, 30m, 35m, 40m, 45m, 50m	c_rng5; c_rng10; c_rng20; c_rng25; c_rng30; c_rng35; c_rng40; c_rng45; c_rng50
Variability of depth across various spatial neighbourhoods	Standard deviation of depths within spatial neighbourhoods equivalent in width to: 5m, 10m, 20m, 25m, 30m, 35m, 40m, 45m, 50m	c_std5; c_ std 10; c_ std 20; c_ std 25; c_ std 30; c_ std 35; c_ std 40; c_ std 45; c_ std 50
Average depth across spatial neighbourhoods	Average of depth within spatial neighbourhoods equivalent in width to: 5m, 10m, 15m, 20m, 25m, 30m, 35m, 40m, 45m, 50m	c_hyp5; c_hyp 10; c_hyp15; c_ hyp 20; c_ hyp 25; c_ hyp 30; c_ hyp 35; c_ hyp 40; c_ hyp 45; c_ hyp 50

Although ocean parameters are also important drivers of benthic community composition and structure (Brown et al. 2011), relevant data were not available for our study areas at spatial and temporal scales sufficiently detailed to justify their use. Further, such data are more readily incorporated into predictive models at biogeographic scales, rather than the regional scale (Williams et al. 2010) implied by the Kimberley offshore region. Acoustic backscatter could also have been useful, as it has shown potential as a benthic biodiversity indicator (McArthur et al. 2010). However, we were not able to use backscatter data for this study for three key reasons. First, collecting such data would have required the vessel to travel at a slower speed, greatly reducing the size of the study areas we could cover across the vast Kimberley region. Second, processing the raw backscatter data collected from the multi-beam is typically computationally expensive while generating data prone to error, particularly artefacts (apparent features that do not actually exist). Finally, the data itself is difficult to interpret ecologically, which meant the costs of collecting it were not justified by its likely value.

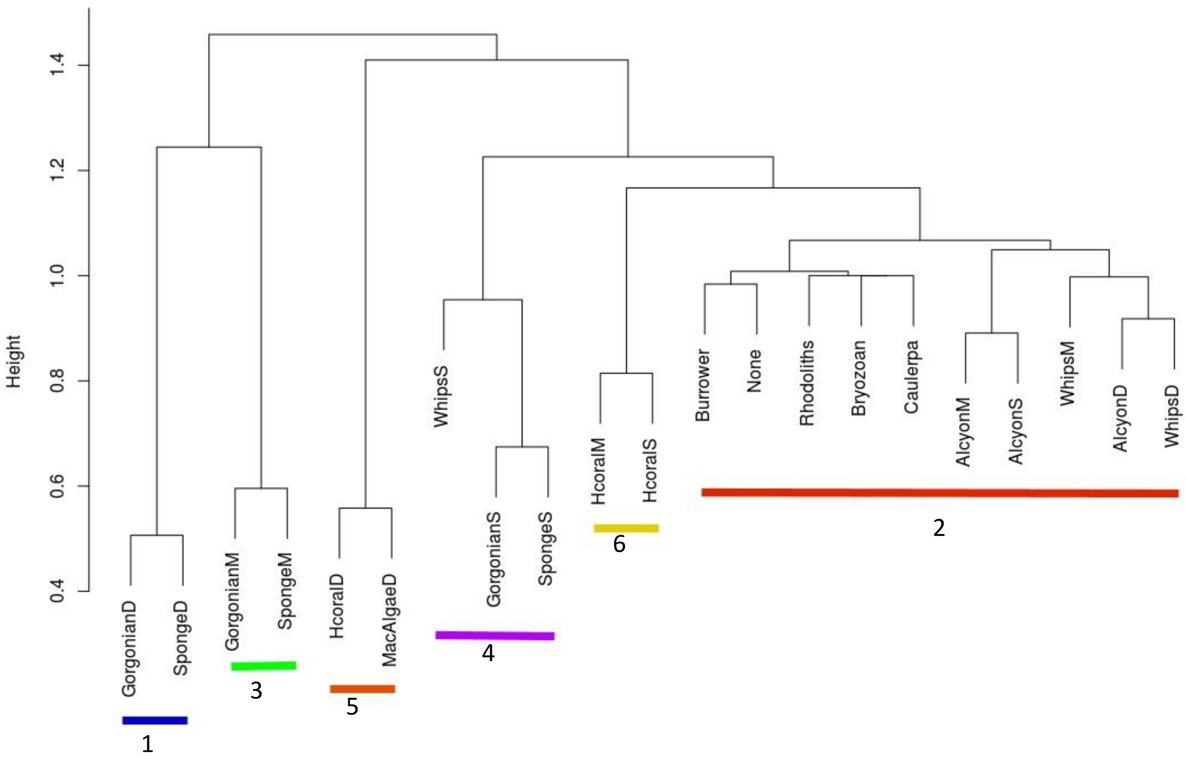
2.2. Developing training and test data

Building predictive models is not possible without verified field data to document where biota of various types actually occur. The data we used to build and test a given model (Figure 2 – steps 3 & 4) came from towed video surveys (Figure 2, step 1) and includes a combination of real-time classification of habitat types from forward-facing video footage and quantitative data from downward facing high-resolution still photos (Figure 6). The location of towed video transects within the study area was determined using a GRTS (Generalized Random Tesselation Stratified) sample design structured to spread transects across *a priori* classes of habitat complexity while ensuring they were evenly distributed spatially (https://science.nature.nps.gov/im/datamgmt/statistics/r/advanced/grts.cfm). The CATAMI classification scheme (http://catami.org/classification) was used to assign benthic categories both for real-time video and for still images.

Figure 7: The AIMS towed video system tow body with mounted video camera and downward facing camera for stills (lower), with an example of a still photo taken just above the sea floor (above).

2.3. Defining mixed classes

From past work in the region, it was apparent that benthic communities often exist in intermixed assemblages which poses considerable challenges for predicting their spatial distribution with high reliability. Where two or more classes typically intermingle in a given study area, a model that attempts to distinguish between them will often confuse these classes with one another. Further, combining probability of occurrence maps for individual classes requires often quite complex ecological and management justifications backed up with existing research and data. In situations where such justification and data are not available, combined models are often based on providing outcomes with the highest accuracy. As a result, decisions about how to label classes when more than one biotic class occurs in a given pixel becomes ecologically ambiguous and can produce misleading results. To avoid these issues and to make the modelling process more ecologically meaningful and robust, we developed a process to create mixed benthic classes prior to modelling (Figure 2, step 5). This was done via cluster analysis of the observed data based on levels of co-occurrence from all three study areas. This determined which individual benthic classes tended to be located within a given distance of one another the most frequently. The process was repeated across a range of spatial scales (10m, 15m, 20m, 25m, 30m, 40m, 50m, 100m) to ensure that the chosen mixed classes were not an artefact of scale. The clusters were then only retained if it seemed ecologically feasible that the grouped classes would have similar requirements and are known to exist in mixed assemblages. The resulting six mixed benthic classes are shown in Figure 7.



1 = Dense Gorgonian, Dense Sponge

2 = Alcyon, Burrowers, Caulerpa, Rhodoliths, Medium Whips, Dense Whips, No benthos.

3 = Medium Gorgonian, Medium Sponge.

4 = Sparse Gorgonian, Sparse Sponge, Sparse Whips.

5 = Dense Hard Coral, Dense Macroalgae.

6 = Sparse Hard Coral, Medium Hard Coral, Bryozoans.

Figure 8: Mixed benthic classes defined based on an 'a priori' cluster analysis (using Euclidian distances and Wards Measure) of the observed data from all three study areas to determine which individual benthic classes were located within 40 m of one another the most frequently.

We withheld a random sample of one-third of the field data to use for model performance estimates (testing set) and used two-thirds of it to establish how benthic classes of organisms vary with the potential predictors (training set) to enable building a model. When establishing the testing and training sets, we also tested for spatial autocorrelation. Where spatial autocorrelation existed, we retained a representative data point for each cluster of auto-correlated points. After each model was built with the training data, we used the testing data to assess its performance (step 8, Figure 2) based on the AUC – 'area under curve' parameter in ROC analysis (Fawcett 2006). Models whose AUC values were less than 0.7 were discarded as per Hosmer et al. (2013).

2.4. Building models and mapping habitats

The statistical relationship between the predictors and testing data was explored using a random forest classification tree model (Figure 2, step 5 – Breiman et al. 1984) which has become a well-established method used in ecological modelling (Cutler et al. 2007, Elith et al. 2006, Elith et al. 2009, Elith et al. 2011) of the marine environment (e.g, marine mammals – Davidson et al. 2012; soft bottom benthic biodiversity - Huang et al. 2014). randomForest provides a method that can fit both linear and non-linear models without overfitting. This method lends itself to ecological interpretation and can produce highly accurate models (Breiman 2001, Cutler et al. 2007). A randomForest model first builds hundreds of classification trees that identify all the unique combinations of variables that could predict the distribution of a given benthic class. Those trees that are not useful in predicting that class cancel each other out. This method outperforms standard classification trees that are defined *a priori* because it ensures that valid relationships in the data are not missed (Cutler et al. 2007).

We built a separate model for each class of benthos that predicts the likelihood of a class existing (Figure 2, step 6) in each pixel across the study area from 0 (no chance it exists) to 1 (100% certainty that it exists). Such maps were generated for each of the six mixed benthic classes, for each of the three study areas (Figure 2 – step 7). In Southern, class 2 was by far the most likely to occur in the study area (Figure 8 – red). This class includes habitats with a variety of taxa, but the abundance is low, typically <1%. Class 2 represents the extensive bare areas with mostly no visible epibenthos, although in places sparse biota are present.

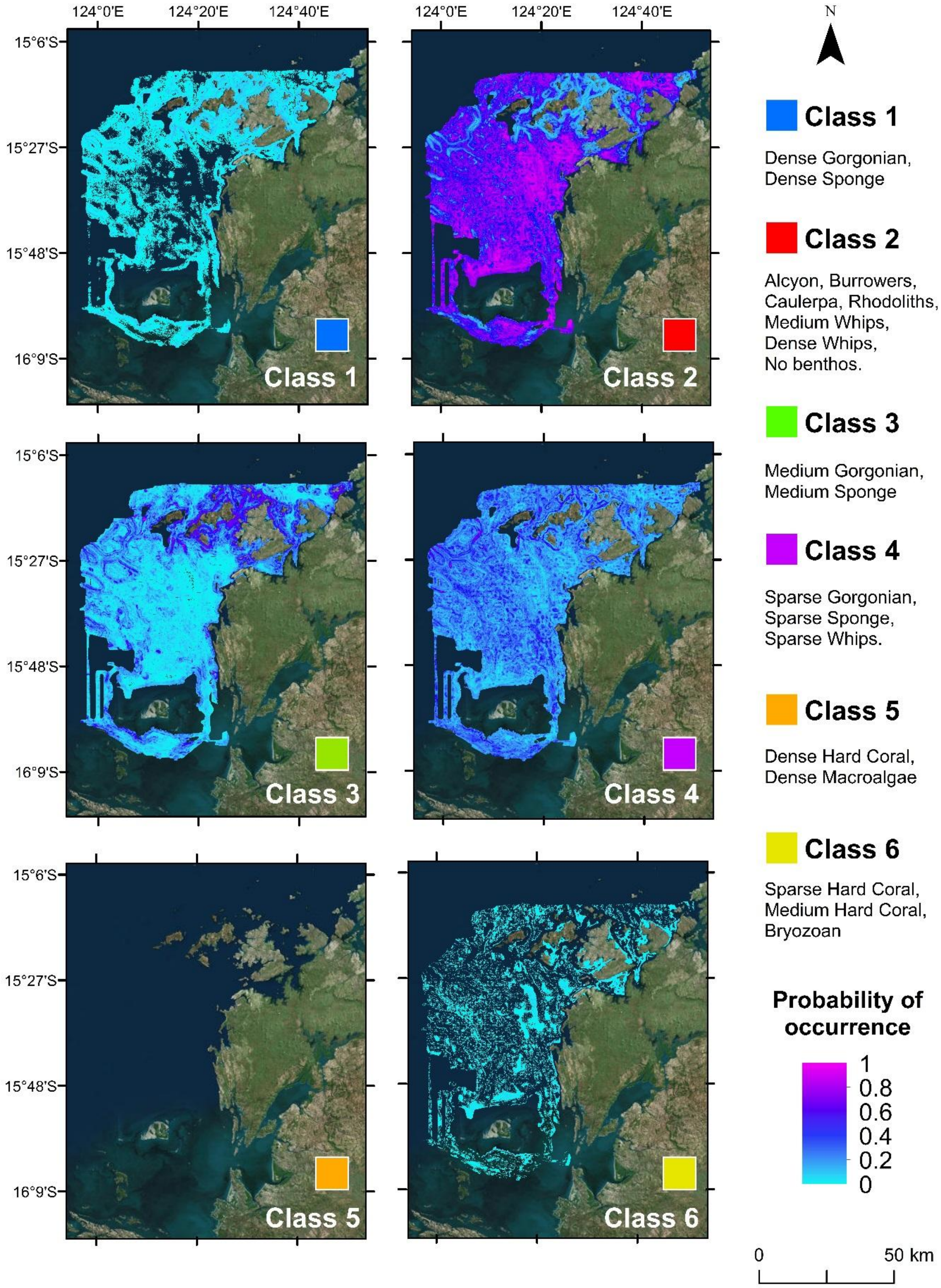


Figure 9: Probability of existence of mixed benthic classes in Southern (campaign 1).

The six most important predictors in the randomForest model for Southern were: depth ['c_depth'], range of depth at local patch ['c_rng10'] and landscape ['c_rng50'] scales, and variability of depth at local ['c_std20'] to moderate ['c_std25', 'c_std30'] scales (Figure 9). Of these, depth was by far the most important.

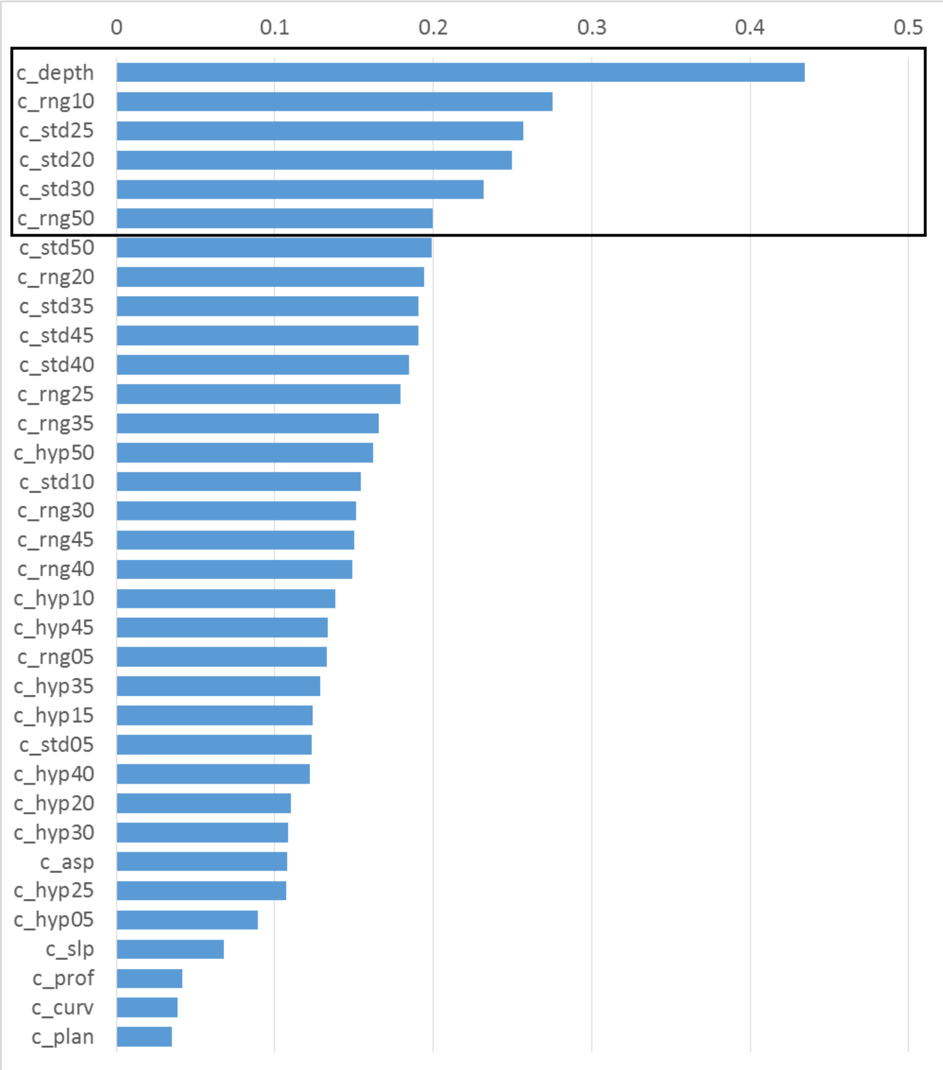


Figure 10: Relative importance of benthic habitat predictor variables in declining order of AUC score for the Southern study area. The names and descriptions of the predictor variables are presented in Table 1.

Figure 10 shows how the likelihood of each class of benthos varied with values of the top six predictors. Class 2 (mixture of Alcyon, Burrowers, Whips and No Benthos – red line) was predicted to be most likely to exist in shallow areas ('c_depth' values between -40 and 0 m) where depth changes very gradually at both very local ('c_rng10' near 0) and moderate ('c_rng50' < 25) scales. Class 2 was also predicted to occur where structural complexity (rugosity) was minimal, as measured at scales from medium to broad (c_std20, c_std25, c_std30, c_std50). In contrast, a general lack of hard substrate across the study area limited the likelihood and spatial distribution of high concentrations of benthos requiring this for attachment – such as Hard Corals (class 5 – orange line- entirely absent; class 6 – yellow line- very unlikely). Class 3 (medium gorgonian and medium sponge – green line) was somewhat likely to occur, particularly in very shallow depths ('c_depth' -20 to 0 m) and where the landscape scale measure of rugosity ('c_rng50' > 35) was moderate to high. Class 4 (sparse gorgonians, sponges and whips – purple line) was slightly more likely to occur then class 3, with it being most prevalent in areas deeper than about 50 m.

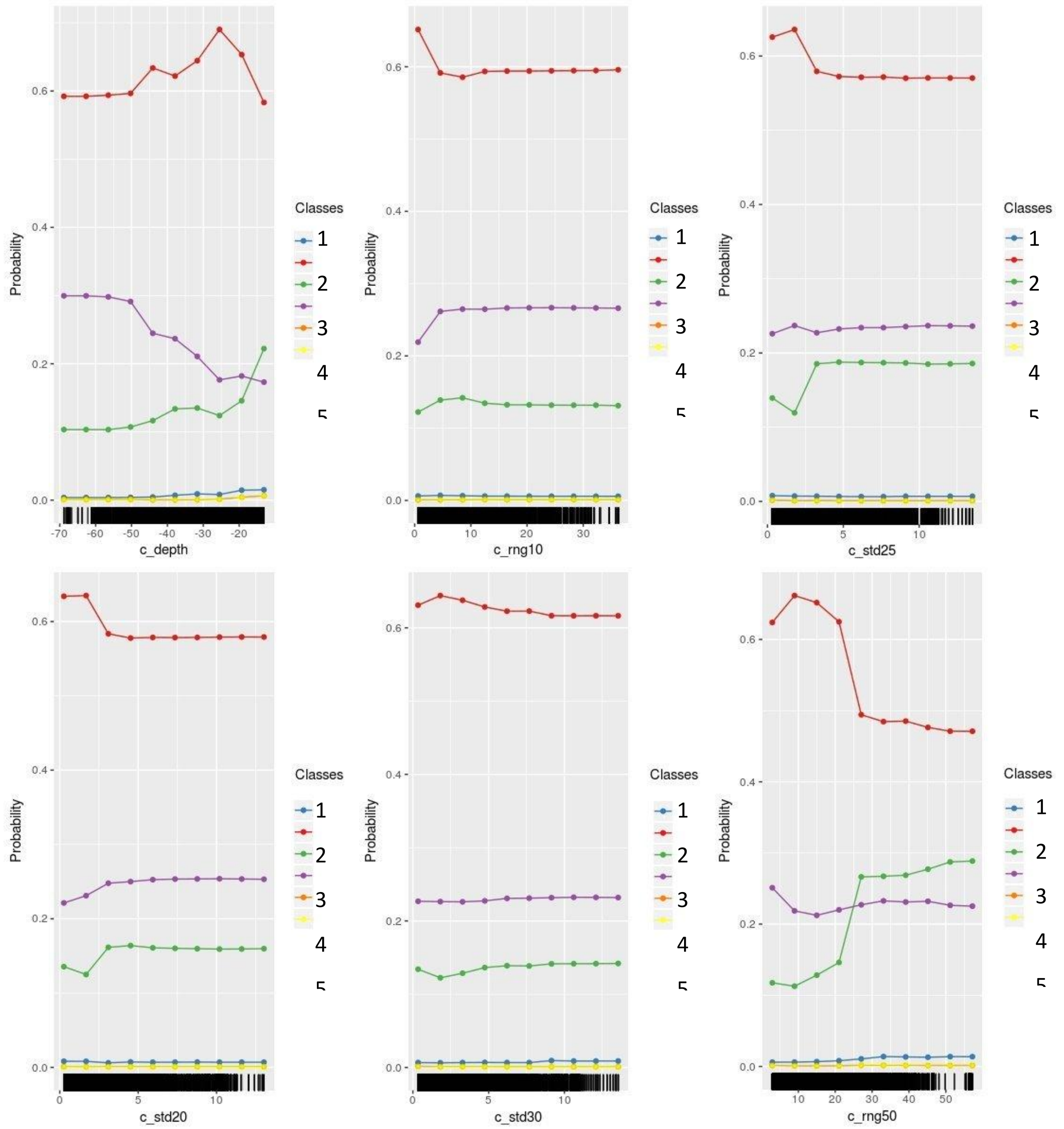


Figure 11: Partial response plot for a priori defined benthic classes for Southern. Mixed benthic classes are as follows: 1 – dense Gorgonian, dense Sponge, 2 – Alcyon, Burrowers, Caulerpa, medium/dense Whips, No Benthos, 3 – medium Gorgonian, medium Sponge, 4 – sparse Gorgonian, sparse Sponge, sparse Whips, 5 – dense Hard coral, dense Macroalgae, 6 – sparse / medium Hard coral, Bryozoans.

A similar pattern is evident for the Central study area (Figure 11), with the dominance of class 2 (red) even more evident than for Southern (Figure 8). Small patches where class 4 (purple), and to a lesser extent, class 3 (green) was likely to occur were scattered throughout the Central study area (Figure 11). Classes containing Hard Coral (class 5 – orange; 6 - yellow) were more likely than in Southern (Figure 8), but still rare (particularly for class 5).

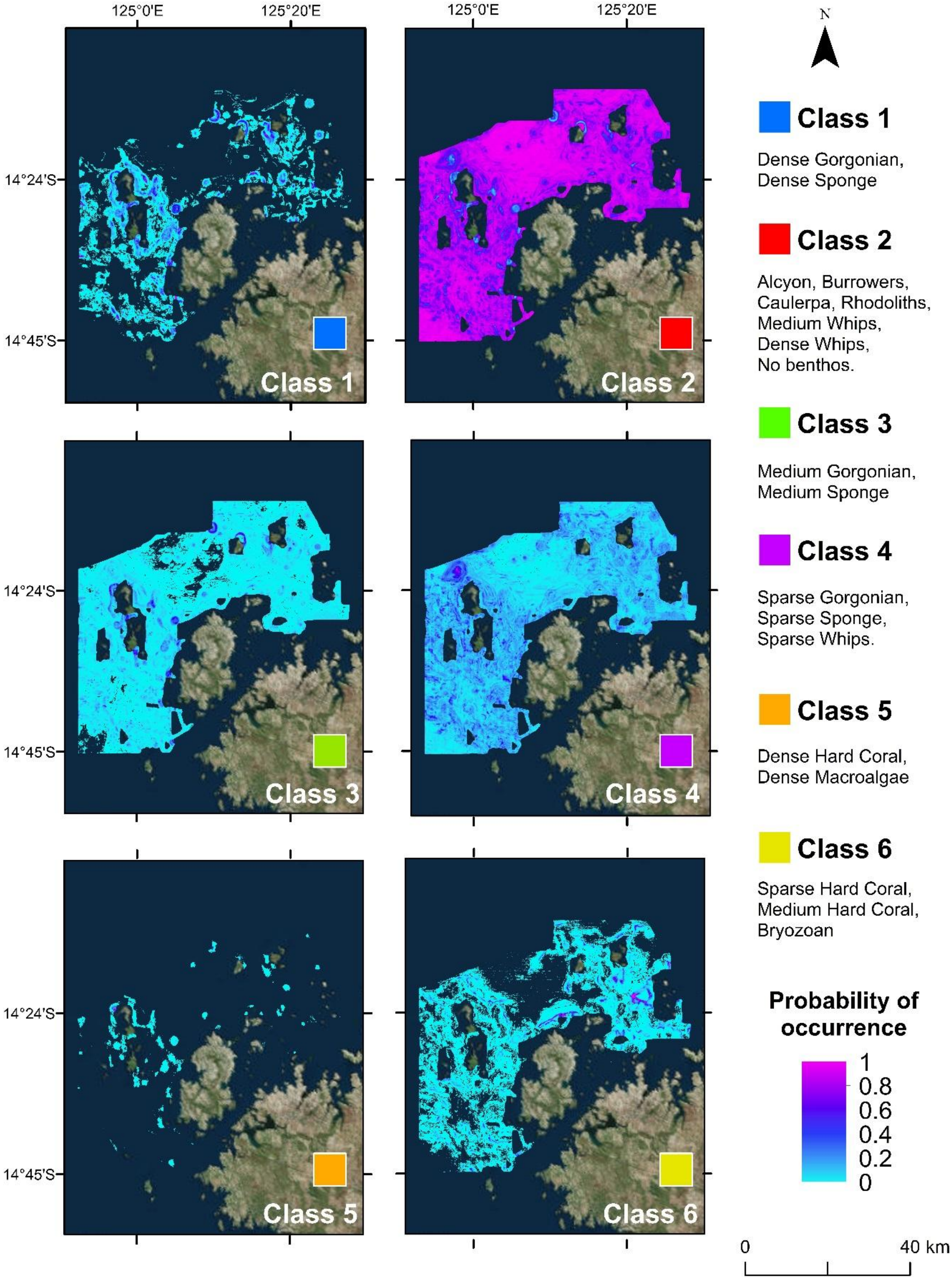


Figure 12: Probability of existence of mixed benthic classes near Central (campaign 2).

The top six predictors in the randomForest model for the Central study area were aspect ['c_asp'], moderate scale variability of depth ['c_std45'], and range of depth at local scales ['c_rng20', 'c_rng10'], depth ['c_depth'], and moderate scales ['c_rng50'] – Figure 12).

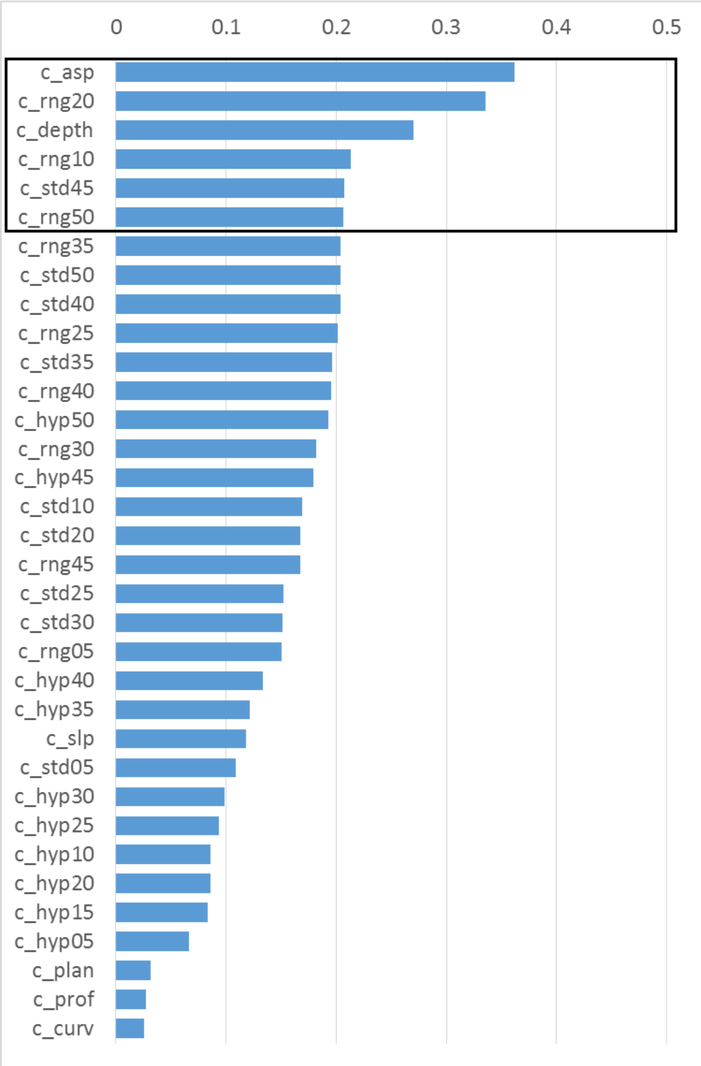


Figure 13: Relative importance of benthic habitat predictor variables in declining order of AUC score for the Central study area. The names and descriptions of the predictor variables are presented in Table 1.

Regardless of predictor values, class 2 (mixture of Alcyon, Burrowers, Whips and No Benthos – red) was always by far the most likely to occur (Figure 13). Class 2 was more likely to exist along easterly facing slopes – 'c_asp' near 100 degrees), while the opposite was true for the other classes. The likelihood of class 2 dropped quickly as local scale measures of range increased (more likely when depth was more uniform), while the opposite occurred for the other classes (more likely when depth was less uniform). For landscape scale measures of range, the response was similar but more gradual. In contrast, the variability of depth at landscapes scales showed the opposite pattern (lower depth variability across larger areas favoured gorgonians and sponges). Finally, class 4 (purple) was the least likely to occur at the deepest depths (~60 m) where class 2 was the most likely to occur. The latter was true to a lesser degree for class 3 (green).

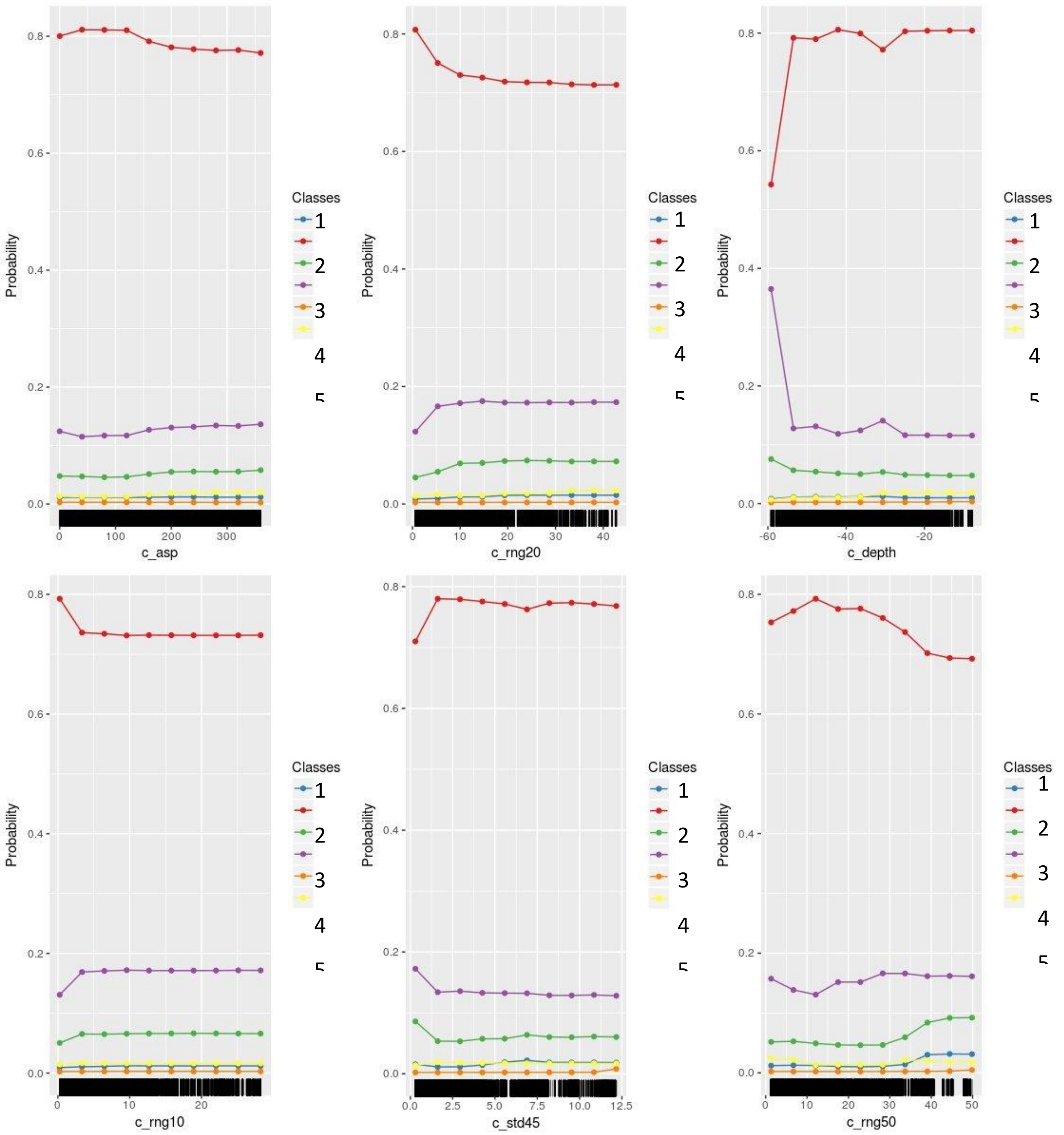


Figure 14: Partial response plot for a priori defined benthic classes for Central. Mixed benthic classes are as follows: 1 – dense Gorgonian, dense Sponge, 2 – Alcyon, Burrowers, Caulerpa, medium/dense Whips, No Benthos, 3 – medium Gorgonian, medium Sponge, 4 – sparse Gorgonian, sparse Sponge, sparse Whips, 5 – dense Hard coral, dense Macroalgae, 6 – sparse / medium Hard coral, Bryozoans.

Hard Coral (5- orange; 6 - yellow) was entirely absent from the Northern study area (Figure 14). Class 2 was again the most likely to exist, followed by class 4 (purple), and to a much lesser extent –class 3 (green). Class 1 (blue) was predicted in narrow, isolated bands where sparse (class 4 - purple) and medium (class 3- green) densities of similar biota were also predicted to occur.

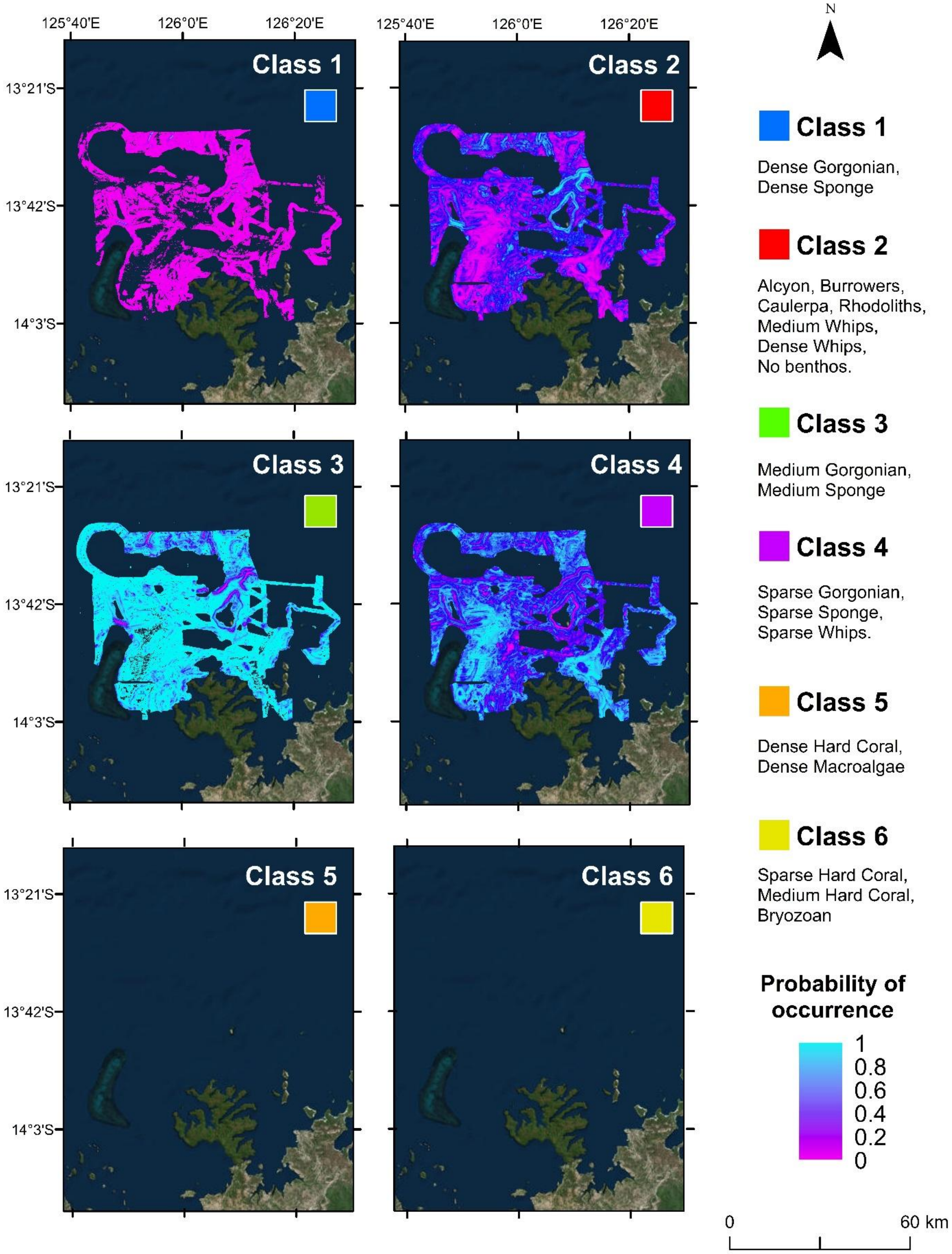


Figure 15: Probability of existence of mixed benthic classes in Northern (campaign 3).

The top six predictors in the randomForest model for the Northern study area were: depth ['c_depth']; variability of depth within landscape scale ['c_std50'], and range of depth within very local ['c_rng10', 'c_rng25'] to landscape ['c_rng45', 'c_rng50'] scales (Figure 15).

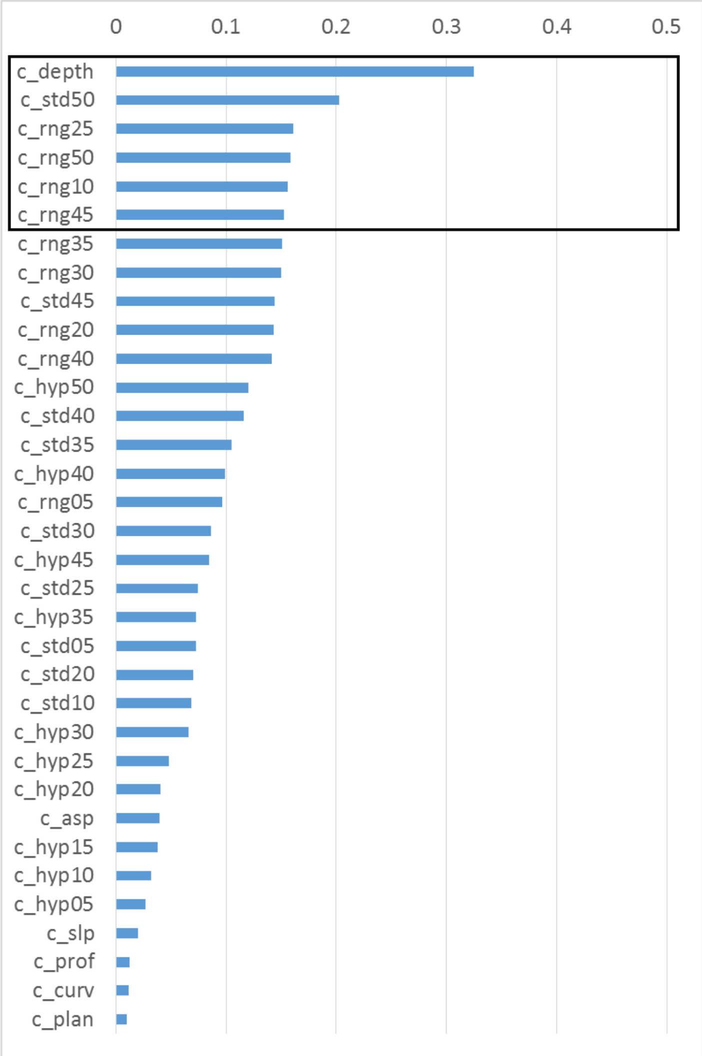


Figure 16: Relative importance of benthic habitat predictor variables in declining order of AUC score for the Northern study area. The names and descriptions of the predictor variables are presented in Table 1.

Regardless of predictor values, class 2 (mixture of Alcyon, Burrowers, Whips and No Benthos – red) was always by far the most likely to occur (Figure 16). Given that, classes 2 and 4 (sparse Gorgonian, Sponge and Whips; purple) showed the strongest response in likelihood to exist with depth, with class 2 more likely to occur in areas shallower than ~45m and class 4 more likely in areas deeper than this. Class 3 (medium Gorgonian and Sponge; green) and class 1 (dense Gorgonian and Sponge; blue) were also slightly more prevalent in deeper waters. More variability in depth at landscape scales (‘c_std50’ > ~7.5) favoured class 2 (red), while the opposite was true for classes 4 (purple) and to a lesser degree 3 (green) and 1 (blue). At all scales, class 2 (red) was more likely to exist when the range of bathymetry was less (more uniform) while the opposite was true for classes 4 (purple), 3 (green) and 1 (blue).

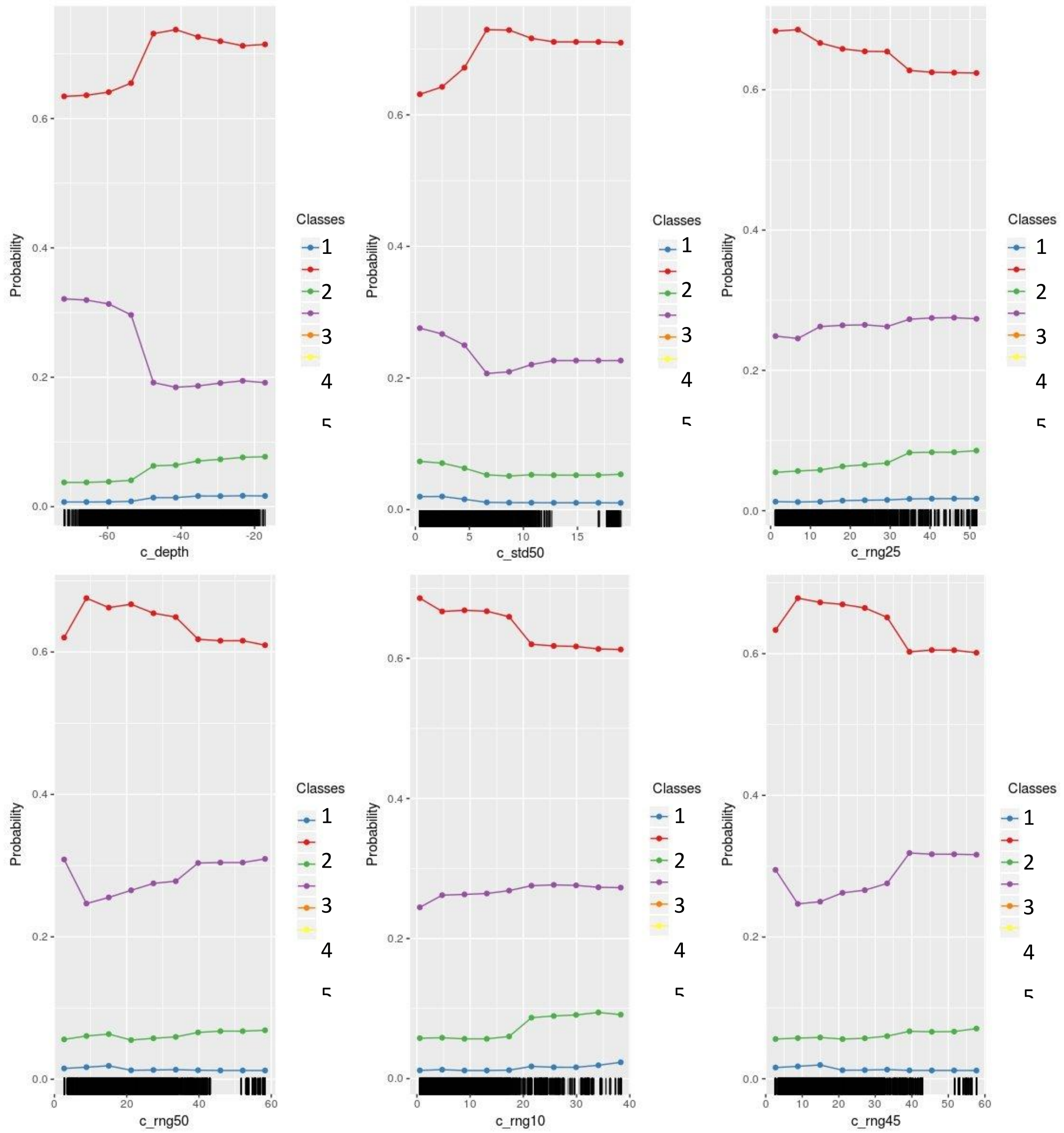


Figure 17: Partial response plot for a priori defined benthic classes for Northern. Mixed benthic classes are as follows: 1 – dense Gorgonian, dense Sponge, 2 – Alcyon, Burrowers, Caulerpa, medium/dense Whips, No Benthos, 3 – medium Gorgonian, medium Sponge, 4 – sparse Gorgonian, sparse Sponge, sparse Whips, 5 – dense Hard coral, dense Macroalgae, 6 – sparse / medium Hard coral, Bryozoans.

2.5. Mapping the most likely mixed class

A maximum likelihood model was used to predict the most likely mixed class in each pixel across the three study areas (Figure 2, step 8) based on the probability of occurrence maps predicted for each mixed class by randomForest (Figure 2, step 7). From this, class 2 was predicted to be by far the most prevalent benthic class across all three study areas (Table 2, Figures 17, 18 and 19 – class 2; red).

Table 2: Percent coverage of each study area by the six mixed benthic classes.

Mixed class code	Mixed class description	% area of Southern (c1)	% area of Central (c2)	% area of Northern (c3)
1	Dense Gorgonian, Dense Sponge	0.03	0.03	0.03
2	Alcyon, Burrowers, Caulerpa, Rhodoliths, Medium Whips, Dense Whips, No benthos	80.08	98.30	92.99
3	Medium Gorgonian, Medium Sponge	12.19	0.81	0.93
4	Sparse Gorgonian, Sparse Sponge, Sparse Whips	7.70	0.78	6.05
5	Dense Hard Coral, Dense Macroalgae	0	0.03	0
6	Sparse Hard Coral, Medium Hard Coral, Bryozoans	0.01	0.06	0

This class is a mixed assemblage dominated by no detectable benthos with scattered patches of Alcyon, Burrowers, medium density Whips and high density Whips, and very rare patches of Caulerpa and Rhodoliths. In contrast, mixed classes containing Hard Coral (classes 5 and 6; orange and yellow) were consistently very rare across the Southern (Figure 8) and Central (Figure 11) study areas and completely absent from the Northern (Figure 14) study area. Class 3 (medium Gorgonian and medium Sponge; green) was predicted to be the most prevalent in the Southern study area (Table 2), particularly in the north-east where it forms large patches around the Augustus and Jungulu islands (Figure 17). Class 3 was predicted to form smaller and much more isolated patches scattered throughout the Central study area (Figure 18), and to form a few long linear patches in the northern section of the Northern study area (Figure 19). Sparse Gorgonian, Sponge and Whips (class 4 – purple) were the least prevalent in the Central study area (Table 2) where they formed compact patches isolated from one another (Figure 18). In contrast, these communities were predicted to exist in long, linear strips near medium Gorgonian and medium Sponge (class 3 – green) in the Southern (Figure 17) and Northern (Figure 19) study areas.

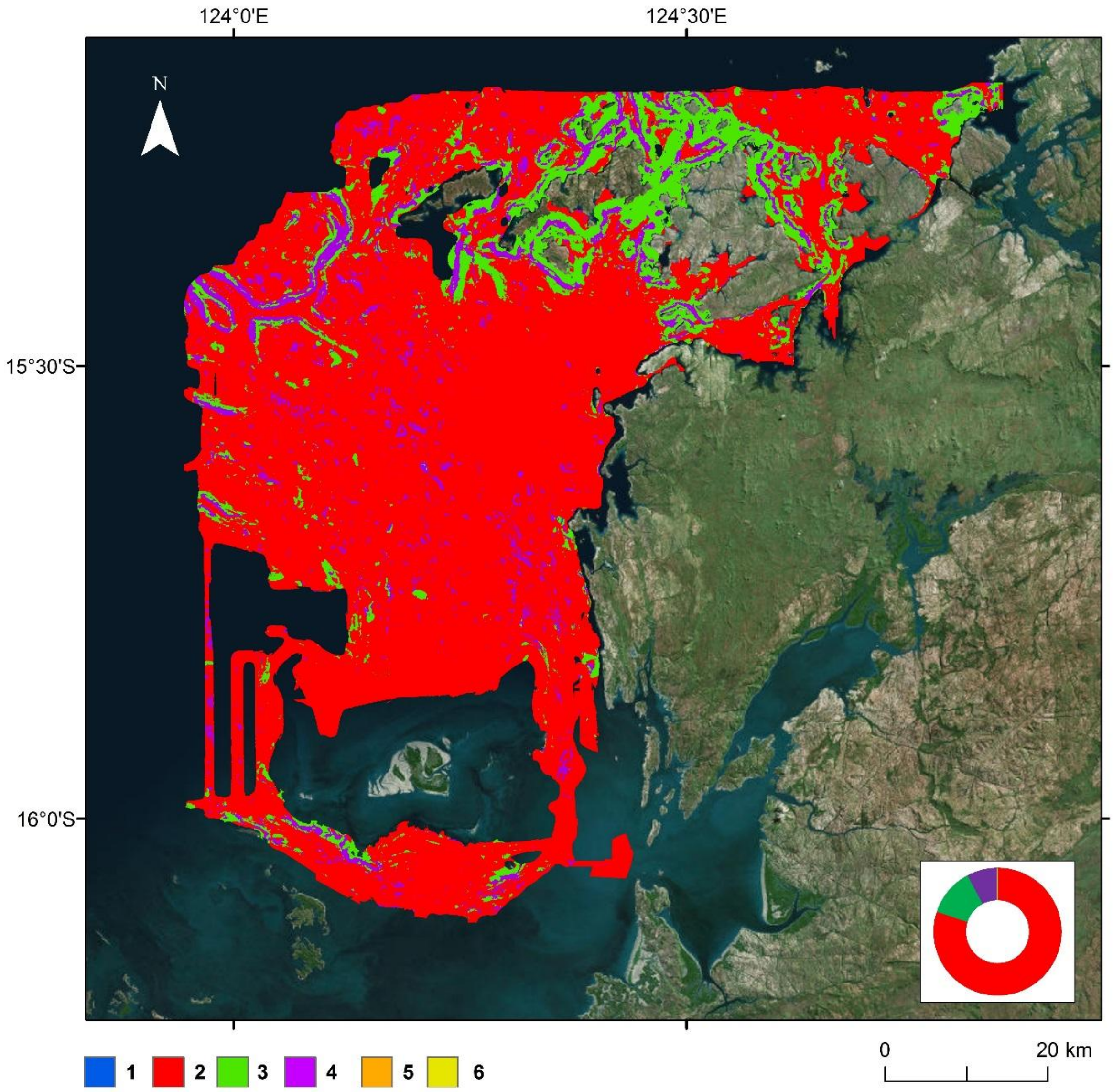
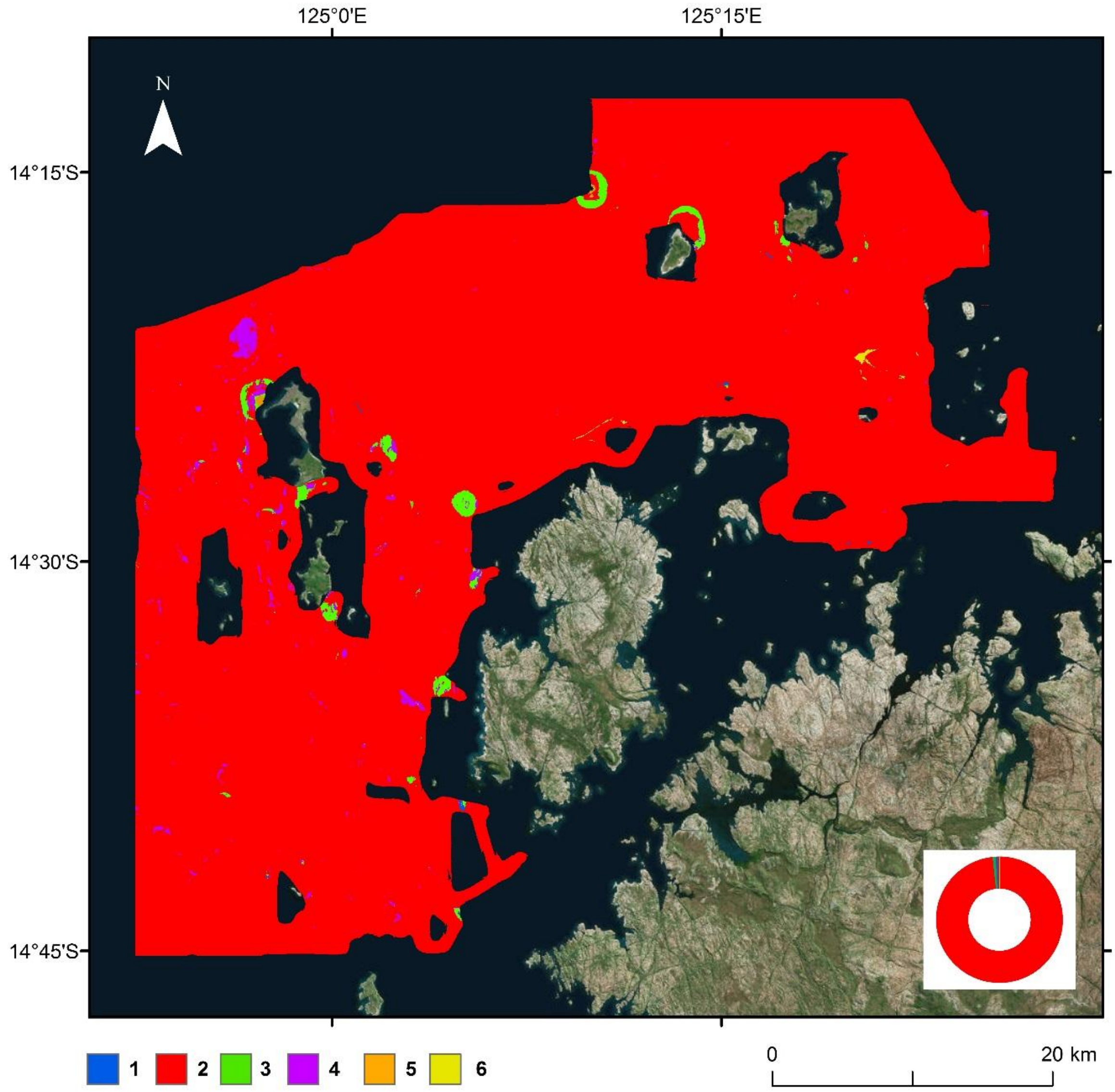


Figure 18: Mixed benthic habitat classification map of the Southern study area (C1). The doughnut diagram shows the % area covered by each mixed class (1 – 0.03%, 2 - 80%, 3 - 12%, 4 - 8%, 5- 0%, 6 – 0.009%). Mixed benthic classes are as follows: 1 – dense Gorgonian, dense Sponge, 2 – Alcyon, Burrowers, Caulerpa, medium/dense Whips, No Benthos, 3 – medium Gorgonian, medium Sponge, 4 – sparse Gorgonian, sparse Sponge, sparse Whips, 5 – dense Hard coral, dense MacroAlgae, 6 – sparse / medium Hard coral, Bryozoans.


125°0'E
125°15'E
N
14°15'S
14°30'S
14°45'S
1
2
3
4
5
6
0
20 km

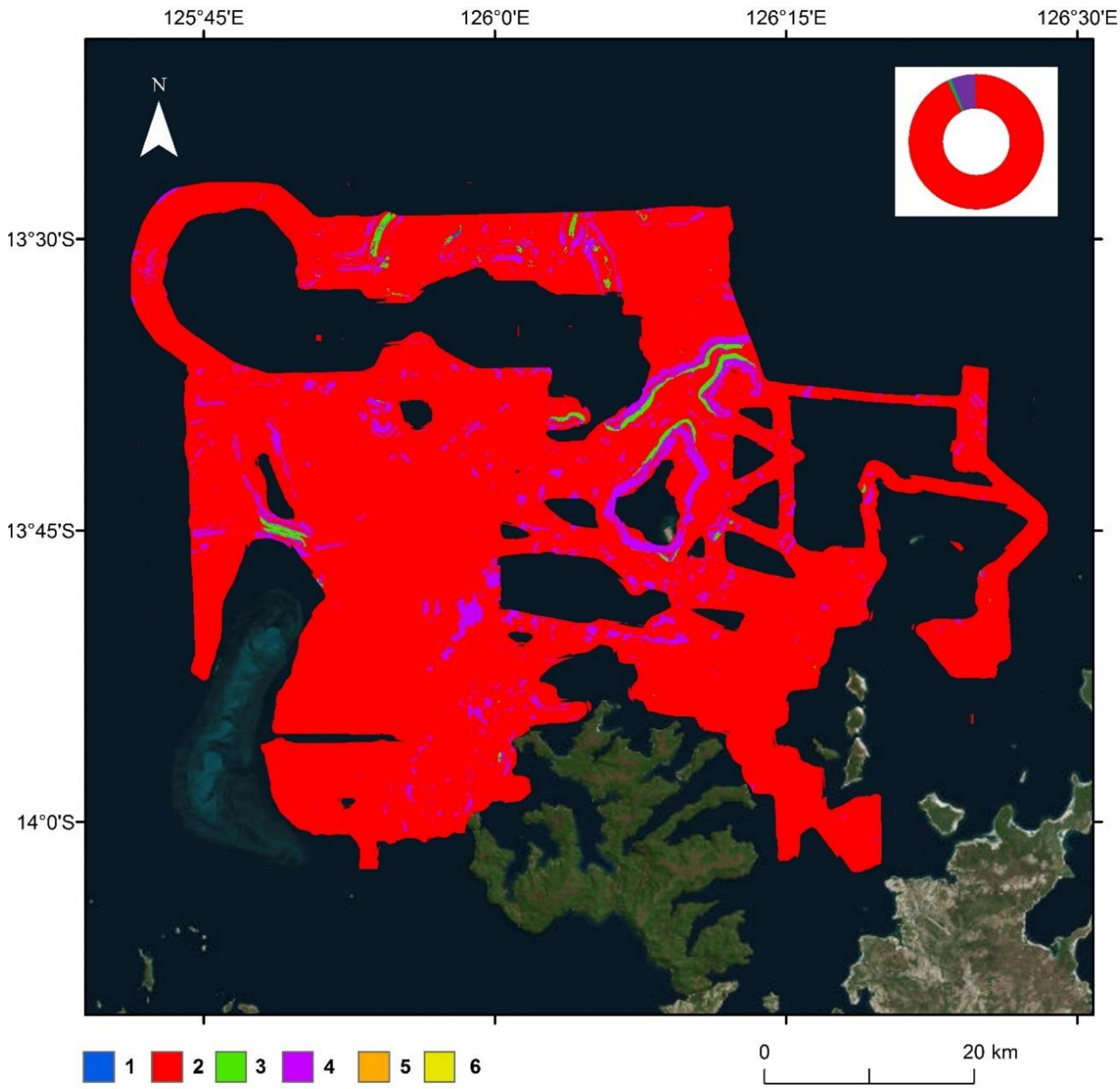


Figure 20: Mixed benthic habitat classification map of the Northern study area (C3). The doughnut diagram shows the % area covered by each mixed class (1 – 0.03%, 2 – 92.99%, 3 – 0.93%, 4 – 6.05%, 5- 0%, 6 – 0%). Mixed benthic classes are as follows: 1 – dense Gorgonian, dense Sponge, 2 – Alcyon, Burrowers, Caulerpa, medium/dense Whips, No Benthos, 3 – medium Gorgonian, medium Sponge, 4 – sparse Gorgonian, sparse Sponge, sparse Whips, 5 – dense Hard coral, dense Macroalgae, 6 – sparse / medium Hard coral, Bryozoans.

2.6. Assessing map accuracy

Once we built a statistical model and used it to predict where a class or classes of biota occurred across the study areas, it was vital to estimate the accuracy of those predictions (Mumby & Harborne 1999; Holmes at al. 2008; Gray 2001). This was done using the testing data points we randomly withheld when building the model (step 4, Figure 2). For each testing data point, we know what actually exists (the observed class), and we know what the model predicts should exist (the predicted class).

Plotting these by benthic class yields what is called a 'confusion matrix' (Congalton 1991). In the confusion matrix, the number of data points where the observed class matches the predicted class is shown for each class in the boxes along the grey shaded diagonal (Figure 20, 21, 22). All the other boxes in the diagram (that are not on the diagonal) indicate misclassification errors – essentially showing all the ways in which the model failed, broken down by class.

	Alcyon, Burrowers, Caulerpa, Rhodoliths, medium Whips, dense Whips, No Benthos	Dense Gorgonian, dense Sponge	Medium Gorgonian, medium Sponge	Sparse Gorgonian, sparse Sponge, sparse Whips	Dense Hard Coral, dense MacroAlgae	Sparse Hard Coral, medium Hard Coral, Bryozoans
Alcyon, Burrowers, Caulerpa, Rhodoliths, medium Whips, dense Whips, No Benthos	**71628**	7	284	5965	0	42
Dense Gorgonian, dense Sponge	27	**569**	113	21	0	0
Medium Gorgonian, medium Sponge	435	163	**12458**	604	0	2
Sparse Gorgonian, sparse Sponge, sparse Whips	2938	7	740	**19500**	0	5
Dense Hard Coral, dense MacroAlgae	0	0	0	0	**0**	0
Sparse Hard Coral, medium Hard Coral, Bryozoans	10	0	9	5	0	**98**
True negative rate (precision)	95.5	76.3	91.6	74.7	-	66.7
True positive rate (recall)	91.9	77.9	91.2	84.1	-	80.3

Figure 21: Confusion matrix plot for mixed classes predicted for the Southern (c1) study area based on sample size = 115,630.

From the confusion matrix, we can calculate how accurate the classification was for each class. The larger the numbers are in the diagonal grey boxes relative to those in the white boxes, the higher the classification accuracy of the model. False negatives occur when the model fails to predict the class that actually exists – these are shown in the white boxes along each *column* in the confusion matrix. For example, in Southern (Figure 20), when the model incorrectly predicted medium Gorgonian / medium Sponge (green) to exist, the actual class was most often sparse Gorgonian / sparse Sponge / sparse Whips (purple). The true positive rate (also known as recall or sensitivity) is measured as 1 minus the rate of false negatives (misses). In this example, it is 91.2%. That means that 91.2% of the time that the class existed, the model predicted it to exist. In other words, the model only missed identifying cases where the class existed (under-predicted) at the data points 8.8% of the time.

	Alcyon, Burrowers, Caulerpa, Rhodoliths, medium Whips, dense Whips, No Benthos	Dense Gorgonian, dense Sponge	Medium Gorgonian, medium Sponge	Sparse Gorgonian, sparse Sponge, sparse Whips	Dense Hard Coral, dense MacroAlgae	Sparse Hard Coral, medium Hard Coral, Bryozoans
Alcyon, Burrowers, Caulerpa, Rhodoliths, medium Whips, dense Whips, No Benthos	**49252**	7	116	727	4	54
Dense Gorgonian, dense Sponge	17	**473**	205	20	0	0
Medium Gorgonian, medium Sponge	128	179	**2454**	315	9	2
Sparse Gorgonian, sparse Sponge, sparse Whips	1221	5	288	**6563**	0	38
Dense Hard Coral, dense MacroAlgae	0	3	10	0	**159**	0
Sparse Hard Coral, medium Hard Coral, Bryozoans	100	0	2	30	0	**935**
True negative rate (precision)	97.1	70.9	79.8	85.7	-	90.9
True positive rate (recall)	98.2	66.2	79.5	80.9	-	87.6

Figure 22: Confusion matrix plot for mixed classes predicted for the Central (c2) study area based on sample size = 63,316.

False positives occur when the model predicts a class, but the observed data tells us something else exists there. These are shown in the white boxes along each *row* in the confusion matrix. For example, in Central (Figure 21), when the observed class was Dense Gorgonian / Dense Sponge (blue), the most common mistake the model made was to predict Medium Gorgonian / Medium Sponge (green). The true negative rate (also known as precision or specificity) is measured as 1 minus the rate of false positives. In this example, it is 70.9%. That means that when the model predicted the class, the class actually existed, just over 70% of the time. In other words, the model wrongly predicted the green class to exist (over-predicted) 29.1% of time.

The true positive and true negative rates can be similar for a given class, as was the case for the benthic classes in the Northern study area (Figure 22). When this happens, the model is just as likely to over-predict a class as it is to under-predict it. In contrast, in Southern (Figure 17), the true negative rate for the sparse Hard Coral / medium Hard Coral / Bryozoan class (yellow) was much higher (80.3%) than the true positive rate (66.7%). That means we can be confident that most of the cases of this class were not missed, but that this came at a cost of predicting the class to exist in some locations where it actually did not exist (over-prediction).

	Alcyon, Burrowers, Caulerpa, Rhodoliths, medium Whips, dense Whips, No Benthos	Dense Gorgonian, dense Sponge	Medium Gorgonian, medium Sponge	Sparse Gorgonian, sparse Sponge, sparse Whips	Dense Hard Coral, dense MacroAlgae	Sparse Hard Coral, medium Hard Coral, Bryozoans
Alcyon, Burrowers, Caulerpa, Rhodoliths, medium Whips, dense Whips, No Benthos	**21494**	59	344	2000	0	0
Dense Gorgonian, dense Sponge	15	**285**	154	22	0	0
Medium Gorgonian, medium Sponge	182	132	**1321**	420	0	0
Sparse Gorgonian, sparse Sponge, sparse Whips	2346	17	365	**6022**	0	0
Dense Hard Coral, dense MacroAlgae	0	0	0	0	**0**	0
Sparse Hard Coral, medium Hard Coral, Bryozoans	0	0	0	0	0	**0**
True negative rate (precision)	89.4	57.8	60.5	71.1	-	-
True positive rate (recall)	89.9	59.9	64.3	68.8	-	-

Figure 23: Confusion matrix plot for mixed classes predicted for the Northern (c3) study area based on sample size = 35,178.

We can also assess accuracy by taking into account how easy it would be to predict a given class purely by chance, which depends in part on the number of data points per class. A widely accepted way to adjust for this is to calculate Kappa (Cohen 1960; noting issues raised in Pontius and Millones 2011):

Kappa = (observed accuracy - expected accuracy)/(1 - expected accuracy)

In essence, Kappa assesses how much better the model predicted the actual classes compared with how well they could be predicted by simply guessing. To interpret the resulting Kappa values, Landis and Koch (1977) rate the skill of the classification into five categories: slight (K=0.01-0.20); fair (K = 0.21-0.40); moderate (0.41-0.60); substantial (0.61-0.80); and almost perfect (0.81-.99). Overall, the classification accuracy of our models as measured by Kappa ranged from substantial to almost perfect for Southern (0.73 to 0.90) and Central (0.68 to 0.92), and from moderate to substantial for Northern (0.55 to 0.68 - Table 4).

Table 3: Classification accuracy estimates (Cohen's Kappa) for each predicted benthic mixed class and overall for each of the three study areas. Landis and Koch (1977) rate the skill of the classification into five categories: slight (K=0.01-0.20); fair (K = 0.21-0.40); moderate (0.41-0.60); substantial (0.61-0.80); and almost perfect (0.81-.99). Missing values occur where none of that class was observed or predicted in a given study area.

Mixed benthic class	Study Area		
	Southern (c1) n = 115,630	Central (c2) n = 63,316	Northern (c3) n = 35,178
Dense Gorgonian, dense Sponge	0.77	0.68	0.58
Alcyon, Burrowers, Caulerpa, Rhodoliths, medium Whips, dense Whips, No Benthos	0.82	0.88	0.68
Medium Gorgonian, medium Sponge	0.90	0.79	0.60
Sparse Gorgonian, sparse Sponge, sparse Whips	0.74	0.81	0.55
Dense Hard Coral, dense MacroAlgae	-	0.92	-
Sparse Hard Coral, medium Hard Coral, Bryozoans	0.73	0.89	-
Across all classes	0.82	0.83	0.65

The highest kappa overall occurred in the Central study area for predicting dense Hard Coral / dense MacroAlgae (orange; K = 0.92) – out of 185 observed locations of this class, only 26 were misclassified by the model (Figure 21). The lowest kappa overall occurred in the Northern study area for predicting sparse Gorgonian / sparse Sponge / sparse Whips (purple; K = 0.55). The latter was driven primarily by confusion with the Alcyon / Burrower / … / No Benthos class (red; Figure 22). Table 4 demonstrates the value of considering how accuracy varies by class – Southern (c1) and Central (c2) scored almost identically for Kappa across all classes (0.82 versus 0.83 – Table 4), yet Kappa differed considerably between them. For example, lower accuracy in classes 4 (purple – Sparse Gorgonian, sparse Sponge, sparse Whips) and 6 (yellow - sparse Hard Coral, medium Hard Coral, Bryozoans) across Southern was balanced by very high accuracy in class 3 (green - medium Gorgonian, medium Sponge). In contrast, across Central, the lowest accuracies were found for classes 1 (blue - dense Gorgonian, dense Sponge) and 3 (green - medium Gorgonian, medium Sponge), and were offset by the highest for class 5 (orange - Dense Hard Coral, dense Macroalgae). By examining the relevant confusion matrices (Figures 20, 21), one can identify the types of mistakes made by the model and assess their importance given the aims of the research.

We tested the impact of using the mixed benthic classes we defined above by calculating the classification accuracy of models based on individual versus those based on mixed classes and comparing them (Table 5).

Table 4: Improvement of classification accuracy (as measured by Kappa – K) with use of mixed instead of individual classes summarised for each of the three study areas. Values of Kappa range from 0 (classification 0% accurate) to 1 (classification 100% accurate). Landis and Koch (1977) rate the skill of the classification into five categories: slight (K=0.01-0.20; fair (K = 0.21-

0.40); moderate (0.41-0.60); substantial (0.61-0.80); and almost perfect (0.81-.99). Kappa values > 0.6 are highlighted in orange and those above 0.8 are highlighted in red.

Study area	Classification accuracy indicator	Model approach	
		individual classes	mixed classes
Southern (c1)	Minimum of K	0.00	0.73
	Maximum of K	0.87	0.90
	Mean of K	0.45	0.79
	Standard deviation of K	0.25	0.06
	% classes where K > .60	21	100
Central (c2)	Minimum of K	0.00	0.68
	Maximum of K	0.97	0.92
	Mean of K	0.57	0.83
	Standard deviation of K	0.23	0.08
	% classes where K > .60	39	100
Northern (c3)	Minimum of K	0.10	0.55
	Maximum of K	0.78	0.68
	Mean of K	0.39	0.60
	Standard deviation of K	0.19	0.05
	% classes where K > .60	11	25

In every case, average classification accuracy was notably higher for the mixed class models than the individual models, nearly doubling for Southern and Northern (Table 5). Further, the standard deviation of Kappa decreased substantially for the mixed class models, indicating that the per-class Kappa rose overall, not just for one or two classes. However, the maximum Kappa value of any class dropped slightly. This suggests that the overall improvements in Kappa across the classes for the mixed models came at the slight expense of the highest performing class in the individual models. For Northern, for example, an individual class of No Benthos was the highest performer (0.78), the performance of which dropped somewhat (0.68) when part of a mixed class.

Classification accuracy per class and over all classes provides an overview of how well the model predicted across an entire study area as a whole (Table 4), but gives no information about how its performance varies within a study area (Comber et al. 2017). In collaboration with Dr Jennifer Miller of the University of Texas at Austin, we mapped how the quality of the classification varied within each study area (Comber et al. 2017; extending methods from Comber et al. 2012, Comber 2013 and Tsutsumida & Comber 2015). Examining the results indicates that high classification accuracy was generally widespread across Southern (more green than purple areas on Figure 23), as implied by the overall high classification accuracy (82%) reported in Table 4.

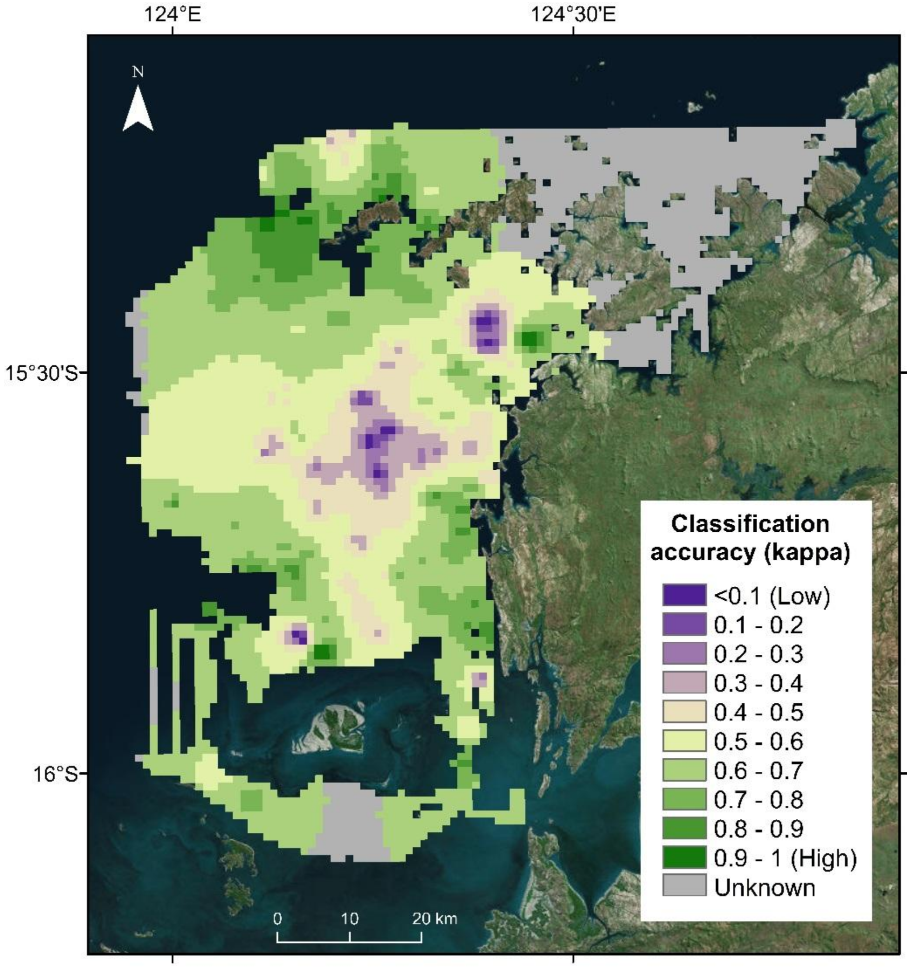


Figure 24: Spatial distribution of classification accuracy (as measured by Cohen's kappa) for the Southern study area (c1). Grey shaded areas were located too distant from sample points to estimate accuracy reliably. Landis and Koch (1977) rate the skill of the classification into five categories: slight (K=0.01-0.20); fair (K = 0.21-0.40); moderate (0.41-0.60); substantial (0.61-0.80); and almost perfect (0.81-.99). On the map, slight to fair performance is shown in purples and moderate to almost perfect performance in greens.

Areas of low classification accuracy (below 0.4; purple areas on Figure 23) are concentrated in pockets of the study area where variation in depth was minimal. This implies less physical structure in the terrain, which suggests the presence of biota that do not require a hard substrate – such as Burrowers. Such terrain may contain isolated hard substrate at very local scales (<1m), supporting scattered patches of sparse biota like Alcyon, Sponges, Whips and Gorgonians. Indeed, the model was in error the most when trying to distinguish between class 2 (red - Alcyon, Burrowers, Caulerpa, Rhodoliths, medium Whips, dense Whips, No Benthos) and class 4 (purple - Sparse Gorgonian, sparse Sponge, sparse Whips) – see Figures 17 and 20.

Though the overall classification accuracy for the Central study area was high (Kappa = 0.83 – Table 4), areas of low accuracy were more spatially widespread (purple areas - Figure 24) than was the case for the Southern study area (Figure 23).

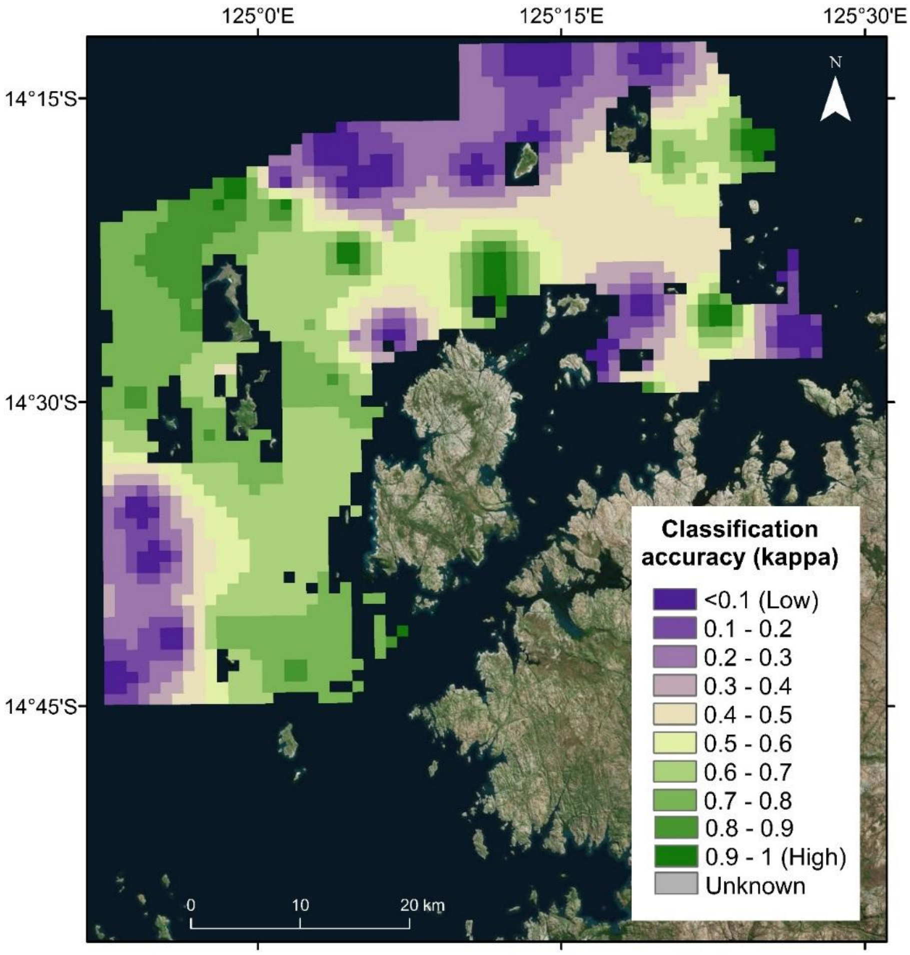


Figure 25: Spatial distribution of classification accuracy (as measured by Cohen's kappa) for the Central study area (c2). Landis and Koch (1977) rate the skill of the classification into five categories: slight (K=0.01-0.20); fair (K = 0.21-0.40); moderate (0.41-0.60); substantial (0.61-0.80); and almost perfect (0.81-.99). On the map, slight to fair performance is shown in purples and moderate to almost perfect performance in greens.

The confusion matrix for the Central study area (Figure 21) shows that the model failed most often when trying to distinguish class 1 (blue - Dense Gorgonian, dense Sponge) from class 3 (green - Medium Gorgonian, medium Sponge), and class 3 (green - Medium Gorgonian, medium Sponge) from classes 4 (purple - Sparse Gorgonian, sparse Sponge, sparse Whips) and 1 (blue - Dense Gorgonian, dense Sponge). This is reflected in the lower Kappa values for these classes in the Central versus Southern study areas (class 1- 0.68 vs 0.77; class 3 – 0.29 vs 0.90; Table 4). This may be due to the relative scarcity of hard substrate in Central versus Southern, making suitable substrate for attachment patchy at very local scales. It could also reflect stochastic recruitment and disturbance processes that have limited the colonisation of habitats otherwise suitable for Gorgonians and into dense communities. And, most importantly, almost twice as many test points were observed across Southern (63,316 versus 115,630 - Table 4).

In contrast, the overall classification accuracy for the Northern study area (Kappa = 0.65 – Table 4) was notably lower than for either Southern or Central, and areas of low accuracy (purple areas) were much more widespread spatially than areas of high accuracy (green areas - Figure 25).

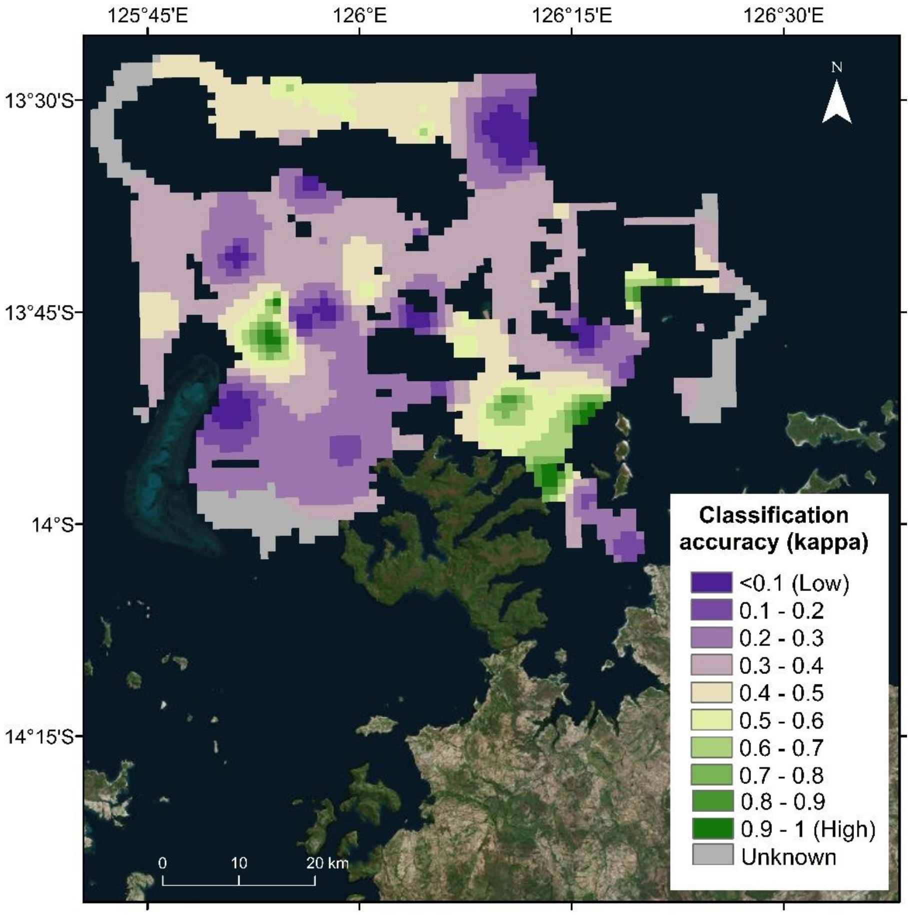


Figure 26: Spatial distribution of classification accuracy (as measured by Cohen's kappa) for the Northern study area (c3). Landis and Koch (1977) rate the skill of the classification into five categories: slight (K=0.01-0.20); fair (K = 0.21-0.40); moderate (0.41-0.60); substantial (0.61-0.80); and almost perfect (0.81-.99). On the map, slight to fair performance is shown in purples and moderate to almost perfect performance in greens.

The confusion matrix for Northern (Figure 22) shows that the model failed most often when trying to distinguish class 1 (blue - Dense Gorgonian, dense Sponge) from class 3 (green - Medium Gorgonian, medium Sponge), and class 3 (green - Medium Gorgonian, medium Sponge) from classes 4 (purple - Sparse Gorgonian, sparse Sponge, sparse Whips) and 2 (red - Alcyon, Burrowers, Caulerpa, Rhodoliths, medium Whips, dense Whips, No Benthos). Again, this may be due to the relative scarcity of hard substrate compared to Southern, making suitable substrate for attachment patchy at very local scales. This was exacerbated by the much lower number of test samples (35,178 versus 115,630 for Southern and 63,316 for Central – Table 4). Fewer samples were obtained for this study area due to an extreme lack of visibility during the second half of the voyage when tidal range increased. Finally, it could also reflect stochastic recruitment and disturbance processes that have limited the colonisation of habitats otherwise suitable for Gorgonians, Whips and Sponges into dense communities.

In summary, higher classification accuracies (green segments, Figure 26) were dramatically more widespread for Southern and Central than Northern, as is evident in Figures 23-25. Though classification accuracy overall was 'almost perfect' for both Southern (82% - Table 5.4) and Central (83% - Table 4), the proportion of the study area that was accurate to that level in Southern (Figure 23) was half that of Central (Figure 24) - 3% versus 6% (dark green segments in Figure 26). However, 18% of the Central versus the Southern study area was classified poorly (dark purple segments in Figure 26). Examining only the global overall classification accuracies or even the confusion matrix would have missed these key differences that are evident in Figures 23 and 24.

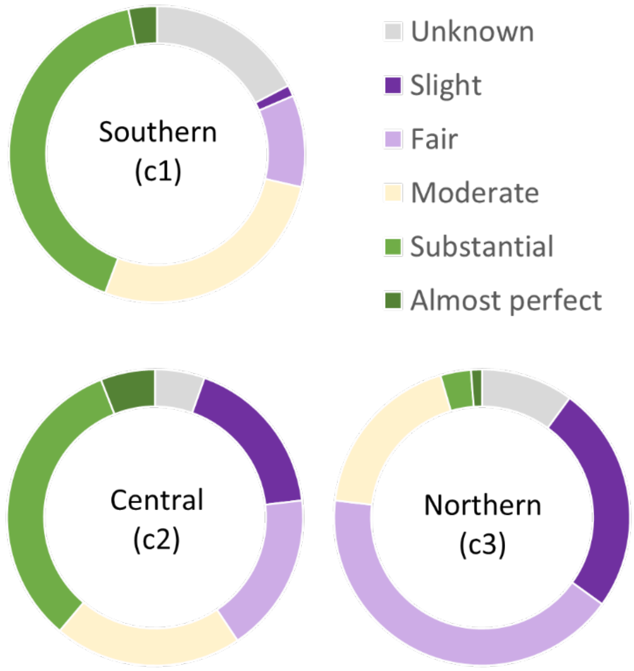


Figure 27: Relative percentage of the Southern (c1), Central (c2) and Northern (c3) study areas for which model classification accuracy was unknown (grey), slight (dark purple - Kappa=0.01-0.20); fair (light purple - Kappa = 0.21-0.40); moderate (beige – Kappa = 0.41-0.60); substantial (light green – Kappa = 0.61-0.80); and almost perfect (dark green – Kappa = 0.81-.99). Classification accuracy was deemed 'unknown' for areas more than 10 km from the nearest test sample point.

We also examined the relative confidence with which the various mixed benthic classes were predicted across the study areas. For example, mixed benthic classes were generally predicted with high accuracy across much of the Southern study area (highest bars in Figure 24 for 'moderate' and 'substantial' Kappa values), and each class was predicted with 'almost perfect' accuracy in some part of the study area. A bit of each benthic class was located too distant from the nearest testing data point to robustly assess classification accuracy – in Southern this represented a notable % area of the study area (high bar for 'Unknown' on Figure 27).

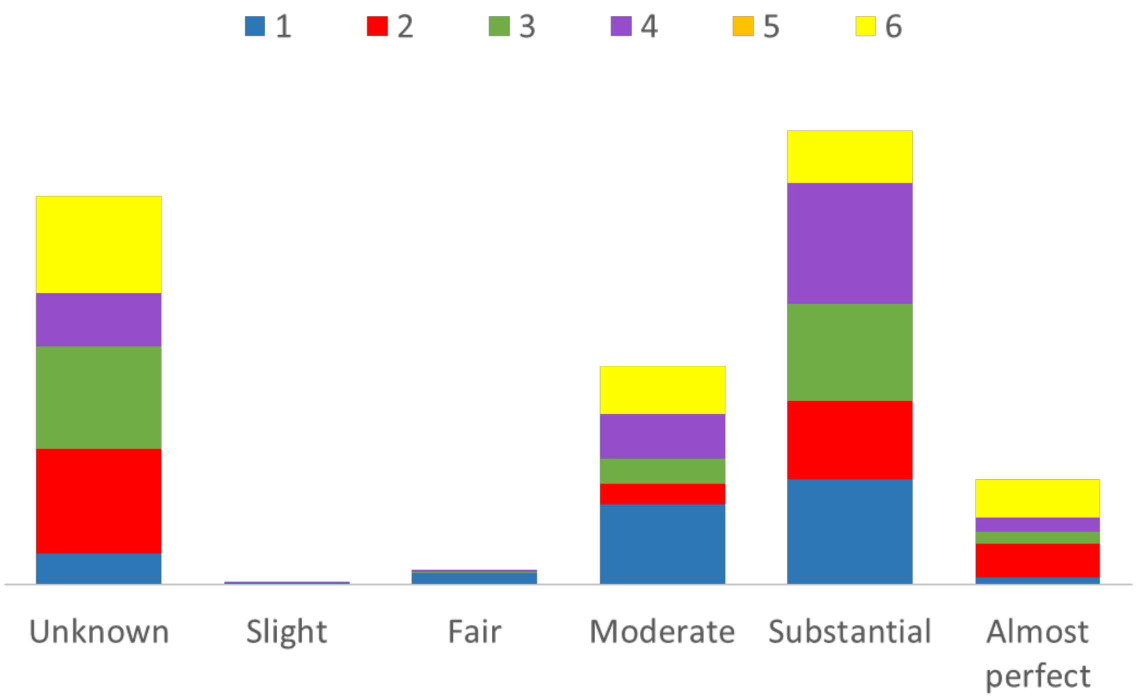


Figure 28: Relative prevalence of each mixed benthic class predicted to occur across Southern (c1), sorted by level of classification accuracy (as measured by Cohen's Kappa). Mixed benthic classes are as follows: 1 – dense Gorgonian, dense Sponge, 2 – Alcyon, Burrowers, Caulerpa, medium/dense Whips, No Benthos, 3 – medium Gorgonian, medium Sponge, 4 – sparse Gorgonian, sparse Sponge, sparse Whips, 5 – dense Hard coral, dense MacroAlgae, 6 – sparse / medium Hard coral, Bryozoans.

In contrast, for Central, a larger proportion of area was classified as 'Almost Perfect' (Figure 28), particularly for classes 4 (purple - Sparse Gorgonian, sparse Sponge, sparse Whips) and 5 (orange - Dense Hard Coral, dense MacroAlgae). However, 'slight' and 'fair' classification accuracy was also much more common than for Southern, particularly for class 1 (blue - Dense Gorgonian, dense Sponge), 3 (green - Medium Gorgonian, medium Sponge) and 5 (orange - Dense Hard Coral, dense Macroalgae). Thus, the poor performance of classes 1, 3 and 5 in some locations balanced its very high performance in others, resulting in an overall classification accuracy very similar to Southern despite very different spatial patterns.

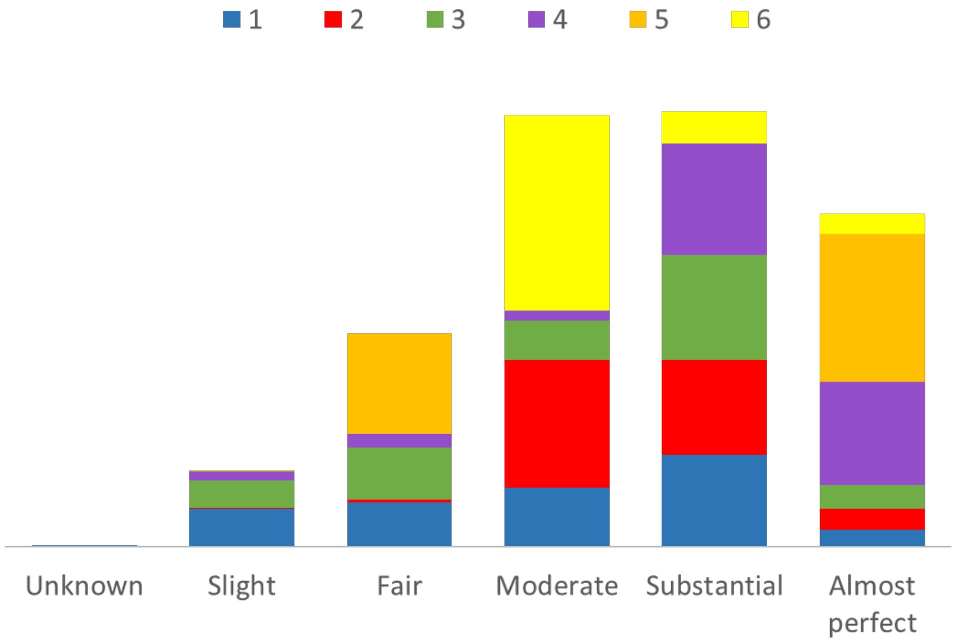


Figure 29: Prevalence of each mixed benthic class predicted to occur across Central (c2), sorted by level of classification accuracy (as measured by Cohen's Kappa). Mixed benthic classes are as follows: 1 – dense Gorgonian, dense Sponge, 2 – Alcyon, Burrowers, Caulerpa, medium/dense Whips, No Benthos, 3 – medium Gorgonian, medium Sponge, 4 – sparse Gorgonian, sparse Sponge, sparse Whips, 5 – dense Hard coral, dense Macroalgae, 6 – sparse / medium Hard coral, Bryozoans.

Lower classification accuracy was much more prevalent across Northern (high bars for 'fair' and 'moderate' in Figure 29) than Southern or Central, though a lower proportion of area fell into the 'slight' category than was the case for Central (Figure 28).

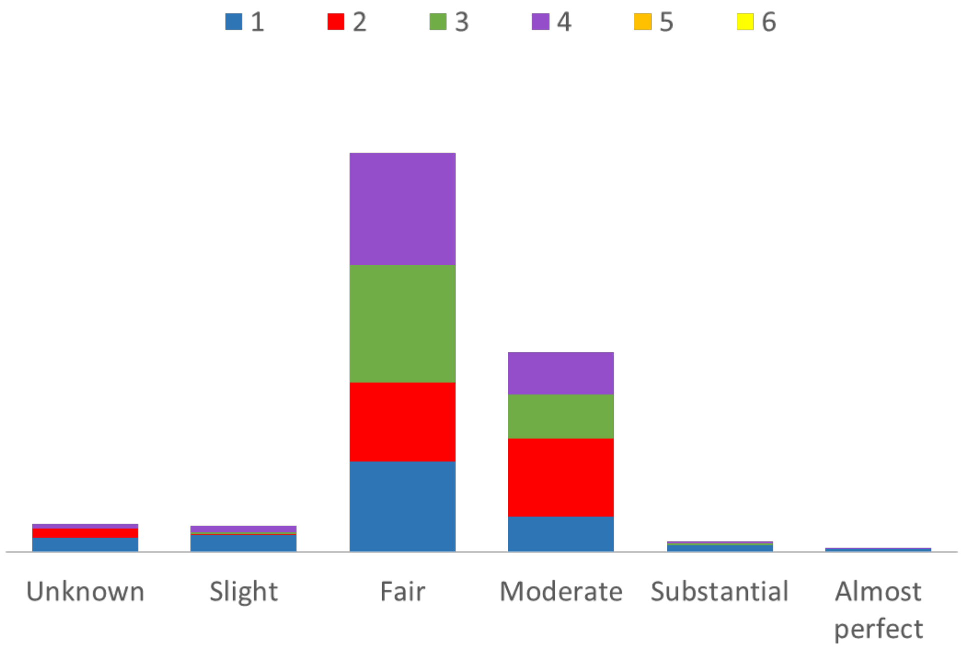


Figure 30: Prevalence of each mixed benthic class predicted to occur across Northern (c3), sorted by level of classification accuracy (as measured by Cohen's Kappa). Mixed benthic classes are as follows: 1 – dense Gorgonian, dense Sponge, 2 – Alcyon, Burrowers, Caulerpa, medium/dense Whips, No Benthos, 3 – medium Gorgonian, medium Sponge, 4 – sparse Gorgonian, sparse Sponge, sparse Whips, 5 – dense Hard coral, dense MacroAlgae, 6 – sparse / medium Hard coral, Bryozoans.

3 Summary of Outcomes

- A mixed benthic class of Alcyon, Burrowers, Caulerpa, Rhodoliths, medium Whips, dense Whips, No Benthos (class 2 – red, Figures 17-19) was predicted to be, by far, the most prevalent class across each of the three study areas (Southern - 80%, Central - 98%, Northern - 93%; Table 3). Within this class, No Benthos and Burrowers were the most prevalent – this makes sense as the majority of the areas surveyed were dominated by sand and mud rather than hard substrate.
- The next most prevalent mixed class was medium Gorgonian and medium Sponge (class 3 – green) for Southern and Central, and sparse Gorgonian, sparse Sponge, sparse Whips (class 4 – purple) for Northern (Table 3). Hard coral was completely absent from the Northern study area at the depths we sampled, and sparse and medium Hard coral was very limited in distribution in the other study areas. This makes sense given that these study areas were dominated by sand and mud with only isolated and very local-scale patches of the hard substrate these communities require for survival.
- We developed a more robust method for dealing with the mixture of predicted classes that typically occur in a given pixel in marine benthic habitat models. In the past, such models either selected the most likely class in a given pixel based on the likelihood of existence of each class, or simplified each probability map to a 'yes/no' map and then created mixed classes based on all the unique combinations of 'yes' values found in a given pixel from all classes. Such a method can miss key biota with limited distribution, especially when pixel sizes are large (low spatial resolution), is time consuming and requires many qualitative decisions throughout the process which limits the reproducibility of the models. By using the observed data to define mixed classes 'a priori' (Figure 2, step 5), we not only reduced computational load, but made the process much more robust (as demonstrated in Table 5), transparent and repeatable.
- In general, validation of the benthic habitat models showed they performed well. Classification accuracy for the study areas overall (Table 4) was 'almost perfect' for Southern (kappa = 0.82) and Central (kappa = 0.83) and 'substantial' for Northern (kappa = 0.65). Examining the spatial distribution of kappa across the study areas showed that areas of high accuracy were most widespread across Southern than for Central (Figure 23, 24, 26), even though their overall kappas were very similar (Table 4). The lower overall classification accuracy at Northern was matched by widespread moderate to low accuracies across that study area (Figure 25).
- The detailed spatial information that mapping classification accuracy provides is vital for making informed spatial allocation decisions for conservation such as determining the placement of protective zones. For example, patches of Dense Hard coral and Dense Macroalgae (class 5 – orange) across Central were predicted with either 'almost perfect' or a lacklustre 'fair' accuracy (Figure 28). Ideally, it would be preferable to protect those patches that were predicted with high rather than low confidence. Our innovative maps of classification accuracy enable such decision processes for the first time.
- This work continues to evolve as we explore this developing research area. Future maps may differ slightly as refinements in methods may occur.

4 References

Breiman, L. (2001), Random Forests, Machine Learning 45(1), 5-32.

Breiman, L (2002), ``Manual On Setting Up, Using, And Understanding Random Forests V3.1'', https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf.

Brown, C. J., Smith, S. J., Lawton, P., & Anderson, J. T. (2011). Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuarine, Coastal and Shelf Science, 92(3), 502-520.

Cohen, J. (1960), "A Coefficient of Agreement for Nominal Scales," Educational and Psychological Measurement, 20, 37-46.

Comber, A., Brunsdon, C., Charlton, M. and Harris, P., 2017. Geographically weighted correspondence matrices for local error reporting and change analyses: mapping the spatial distribution of errors and change. Remote Sensing Letters, 8(3), pp.234-243.

Comber, A. J. (2013). Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies. Remote Sensing Letters, 4(4), 373-380.

Comber, A., Fisher, P., Brunsdon, C., & Khmag, A. (2012). Spatial analysis of remote sensing image classification accuracy. Remote Sensing of Environment, 127, 237-246.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote sensing of environment, 37(1), 35-46.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. Ecology, 88(11), 2783-2792.

Davidson, A. D., Boyer, A. G., Kim, H., Pompa-Mansilla, S., Hamilton, M. J., Costa, D. P., ... & Brown, J. H. (2012). Drivers and hotspots of extinction risk in marine mammals. Proceedings of the National Academy of Sciences, 109(9), 3395-3400.

Edgar, G. J., Bates, A. E., Bird, T. J., Jones, A. H., Kininmonth, S., Stuart-Smith, R. D., & Webb, T. J. (2016). New approaches to marine conservation through the scaling up of ecological data. Annual Review of Marine Science, 8, 435-461.

Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27, 861-374.

Gray, J. S. (2001). "Antarctic marine benthic biodiversity in a world-wide latitudinal context." Polar Biology 24(9): 1432-2056.

Holmes, K. W., et al. (2008). "Modelling distribution of marine benthos from hydroacoustics and underwater video." Continental Shelf Research 28(14): 1800-1810.

Hosmer, D. W. Jr., Lemeshow, S., and Sturdivant, R. X. (2013). Assessing the Fit of the Model. Applied Logistic Regression, 3rd Edn. John Wiley & Sons. doi: 10.1002/9781118548387.ch5

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159-174

Lehmann, A., et al. (2002). "Regression models for spatial prediction: their role for biodiversity and conservation." Biodiversity and Conservation 11: 2085-2092.

McArthur, M. A., Brooke, B. P., Przeslawski, R., Ryan, D. A., Lucieer,V. L., Nichol, S., McCallum, A. W.,Mellin, C., Cresswell, I. D., and Radke, L. C. (2010). On the use of abiotic surrogates to describe marine benthic biodiversity. Estuarine, Coastal and Shelf Science 88, 21–32. doi:10.1016/J.ECSS.2010.03.003

Monk, J., Ierodiaconou, D., Bellgrove, A., Harvey, E., & Laurenson, L. (2011). Remotely sensed hydroacoustics and observation data for predicting fish habitat suitability. Continental Shelf Research, 31(2), S17-S27.

Mumby, P. J. and A. R. Harborne (1999). "Development of a systematic classification scheme of marine habitats to facilitate regional management and mapping of Caribbean coral reefs." Biological Conservation 88(2): 155-163.

Pontius, R G Jr and Millones, M (2011) Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment, International Journal of Remote Sensing, 32:15, 4407-4429, DOI: 10.1080/01431161.2011.552923

Tsutsumida, N., & Comber, A. J. (2015). Measures of spatio-temporal accuracy for time series land cover data. International Journal of Applied Earth Observation and Geoinformation, 41, 46-55.

Williams, A., F. Althaus, P.K. Dunstan, G.C.B. Poore, N.J. Bax, R.J. Kloser and F.R. McEnnulty. (2010). Scales of habitat heterogeneity and megabenthos biodiversity on an extensive Australian continental margin (100-1,000m depths). Marine Ecology 31: 222-236.