Ministry for Primary Industries
Manatū Ahu Matua

Developing predictive models for the distribution of vulnerable marine ecosystems in the South Pacific Ocean region

New Zealand Aquatic Environment and Biodiversity Report No. 120.

A.A. Rowden,
J.M. Guinotte,
S.J. Baird,
D.M. Tracey
K.A. Mackay,
S. Wadhwa

ISSN 1179-6480 (online)
ISBN 978-0-478-42334-1 (online)
December 2013

New Zealand Government

Growing and Protecting New Zealand

Requests for further copies should be directed to:

Publications Logistics Officer
Ministry for Primary Industries
PO Box 2526
WELLINGTON 6140

Email: brand@mpi.govt.nz
Telephone: 0800 00 83 33
Facsimile: 04-894 0300

This publication is also available on the Ministry for Primary Industries websites at:
http://www.mpi.govt.nz/news-resources/publications.aspx
http://fs.fish.govt.nz go to Document library/Research reports

© Crown Copyright - Ministry for Primary Industries

TABLE OF CONTENTS

Executive Summary

Rowden, A.A.; Guinotte, J.M.; Baird, S.J.; Tracey, D.M.; Mackay, K.A.; Wadhwa, S. (2013). Predictive modelling of the distribution of vulnerable marine ecosystems in the South Pacific Ocean region.

***New Zealand Aquatic Environment and Biodiversity Report No. 120.* 70 p.**

Vulnerable marine ecosystems (VMEs) are any deep-sea ecosystems that are highly vulnerable to one or more kinds of fishing activity, and are identified by the vulnerability of their components (e.g. habitats, communities or species). The South Pacific Regional Fishery Management Organisation (SPRFMO) Convention includes specific provisions to protect VMEs.

There are relatively few data available on the distribution of VME indicator species or taxa in the South Pacific Ocean to use for the objective planning of spatial protection measures to protect those taxa, particularly in the SPRFMO area. It is therefore becoming increasingly important to develop robust predictions of where VMEs are likely to occur, using habitat suitability models.

In order to address SPFRMO's specific requirements for robust and reliable models of the distribution of VME indicator taxa in the South Pacific Ocean, a multi-phased research project was proposed to develop predictive models for VMEs in the South Pacific Ocean, and to use these models, in conjunction with appropriate spatial management planning tools, to identify areas of the South Pacific that most warrant protection from the impacts of fishing activities.

The objective of Phase 1 of the project presented here was to collate data sets and evaluate modelling approaches which are likely to be useful to predict the distribution of VMEs in the South Pacific Ocean region.

Data for ten VME indicator taxa were obtained from database sources in New Zealand and overseas. The final dataset contained over 31 000 individual records for Actinaria, Alyconacea, Antipatharia, Brisingida, Crinoidea, Gorgonacea, Pennatulacea, Porifera, Scleractinia, and Stylasteridae.

Data for environmental variables were obtained from various sources. The variables were grouped into two environmental data sets: a 'regional' set of 11 environmental data layers that included variables developed or tuned specifically for the New Zealand region plus a limited number obtained from other mostly regional sources), and a 'global' set of 9 environmental data layers collated and processed from global data sources.

Three types of habitat suitability model were made for each VME indicator taxon: a maximum entropy (Maxent) and a boosted regression tree (BRT) model using the 'regional' environmental data, and a Maxent model using the 'global' environmental data. Model performance was evaluated using model-specific measures and the area under the curve (AUC) measure, a quantitative performance measure common to all model types.

Both Maxent and BRT models perform well. The distribution maps of predicted habitat suitability produced by the Maxent and BRT models for the ten VME indicator taxa were almost all qualitatively dissimilar to one another. In general the Maxent predictions of suitable habitat were constrained to areas where data records exist, while BRT predictions of suitable habitat extended more into unsampled areas. BRT models also produced greater variability in their predictions of habitat suitability at the smallest spatial scales than Maxent models.

The comparison of the Maxent models using 'regional' and 'global' environmental data layers indicated that these two models perform equally well. The distribution maps of predicted habitat

suitability produced by the Maxent models for the ten VME indicator taxa based on the ‘regional’ and ‘global’ environmental data layers were overall qualitatively similar.

All the BRT and Maxent models predicted that large areas of the study region represented suitable habitat for VME indicator taxa. The general inability of models to finely discriminate highly suitable habitat is in part a consequence of the coarse taxonomic levels used to classify VME indicator taxa.

The total number of collated records varied considerably among the different VME indicator taxa, and the number of records that could be used for modelling purposes ranged from 505 to 9187. While there were sufficient records to produce useful models, the distribution of data records across environmental space reflects a sampling bias. Sampling that generated records for VME indicator taxa data used in the modelling has been mainly concentrated in the New Zealand EEZ, with very little sampling conducted in the high seas portion (i.e. the SPRFMO area) of the study area.

The results of Phase 1 of the South Pacific VME Project provide recommendations that can inform the future phases of the project:

- BRT and Maxent, as well as other modelling approaches, should continue to be used;
- the criteria for identifying VME indicator taxa should be revised;
- in order to refine the identification of VMEs, predictive models should be generated that combine the presence of particular taxa;
- a biodiversity survey should be undertaken of the Louisville Seamount Chain (and/or West Norfolk Ridge, Lord Howe Rise) to obtain data for VME indicator taxa and habitat, ground truth the preliminary models, and develop new models;
- new models for the SPRFMO region should include those that model VME habitat (e.g. deepwater coral reef) directly, and if possible incorporate estimates of genetic connectivity, as well as spatially explicit measures of uncertainty;
- where multibeam surveys have been undertaken in the SPRFMO region, bathymetric and backscatter data should be used to make high resolution habitat suitability maps.

1. INTRODUCTION

1.1 Overview

The Convention and Final Act of the South Pacific Regional Fisheries Management Organisation (SPRFMO) were finalised in November 2009, establishing SPRFMO as the international fisheries management organisation with the mandate and responsibility to manage all non-highly migratory marine resources within the high-seas areas of the South Pacific Ocean (www.southpacificrfmo.org). The SPRFMO Convention includes specific provisions to protect vulnerable marine ecosystems (VMEs). A VME is any deep-sea ecosystem that is highly vulnerable to one of more kinds of fishing activity, and it is identified by the vulnerability of its components (e.g. habitats, communities or species) (FAO 2009).

Criteria for identifying VMEs include uniqueness or rarity of species or habitats, their functional significance, fragility, and structural complexity as well as life histories that limit the probability of recovery. The FAO guidelines also provide examples of taxa indicative of a VME: (i) cold-water corals of various types (e.g. reef builders and coral forest species) likely to be found on the edges and slopes of oceanic islands, continental shelves, seamounts, canyons, and trenches; (ii) sponge-dominated communities and structural biogenic habitats (e.g. those composed of large protozoans, hydrozoans or bryozoans) with a distribution similar to cold-water corals; (iii) endemic or rare types of hydrothermal vent and cold seep communities; and (iv) fish species that sustain low exploitation rates (FAO 2009).

SPRFMO has adopted a series of interim measures designed to meet the obligations established by United Nations General Assembly Resolution 61/105, and reinforced in 2009 in resolution 64/72, to prevent ‘significant adverse impacts’ by bottom fisheries on VMEs, where ‘known to occur or are likely to occur’. These interim measures include: constraining fishing effort to previously fished areas; a requirement to implement conservation and management actions to prevent significant adverse impacts on VMEs within these fished areas, and the implementation of a ‘move-on’ rule requiring vessels to move away from areas in which evidence of a VME is encountered during fishing operations (Penney et al. 2009).

The efficacy of move-on rules as an effective measure to prevent impacts on VMEs is increasingly being questioned (e.g. Auster et al. 2011) because of increasing evidence that low trawling effort levels can have significant and long-lasting adverse effects on benthic biodiversity and biogenic habitat structure (Williams et al. 2010), and the risk that such measures will result in the spread of the impacts of trawling. Over the past few years there has therefore been a rapid increase in emphasis on the importance of implementing adequate and representative spatial closures or gear restrictions to protect areas known or likely to support VMEs (Auster et al. 2011, Morato et al. 2010).

There are, as yet, relatively few actual data available on the distribution of VME indicator species or taxa in the South Pacific Ocean to use for the objective planning of spatial protection measures to protect these taxa, particularly in the SPRFMO area. It is therefore becoming increasingly important to develop robust predictions of where VMEs are likely to occur, using habitat suitability models (sometimes referred to as species distribution models). Such models have recently been developed and are in the process of being refined for certain VME indicator taxa on a global scale (e.g. Actinaria, Guinotte et al. 2006; Scleractinia, Tittensor et al. 2009; Octocorallia, Yesson et al. 2012). However, the spatial resolution of existing models is coarse (larger than the scale of the topographic features typically the target of fishing on the high seas), and the level of uncertainty around the predictions is variable or still unknown.

The lack of appropriate data and distribution models for VME indicator taxa was highlighted in 2010 by the Ocean and Marine Sciences workshop of the USA and New Zealand Joint Commission Meeting (JCM), and identified as an area for future collaborative research. There is therefore a need to develop and validate finer-scale, regionally focused models of a greater number of VME indicator taxa, together with attached estimates of model uncertainty, as a first step towards identifying areas likely to contain a high diversity of VMEs, to guide SPRFMO participants in developing measures to protect such areas from significant adverse impacts.

1.2 SPRFMO Requirements

The FAO International Guidelines for the Management of Deep-Sea Fisheries in the High Seas (FAO 2009, paragraph 45) note that, "where site-specific information is lacking, other information that is relevant to inferring the likely presence of vulnerable populations, communities and habitats should be used". To provide guidance to SPRFMO participants on the drafting of bottom fishery impact assessments (which are required in terms of the SPRFMO Interim Measures before any bottom fishing takes place in the SPRFMO Area), the SPRFMO Scientific Working Group has recently (February 2012) developed a 'Bottom Fishery Impact Assessment Standard' (http://www.southpacificrfmo.org/benthic-impact-assessments). The standard, under the section on *Predictors to Evaluate Likelihood of Occurrence of VMEs*, notes that:

"For much of the SPRFMO Area, data on seabed biodiversity and benthic community composition are not available. Therefore, ancillary information on other factors that influence the location of VMEs, will need to be used to predict likelihood and suitability of areas for supporting VMEs."

An example of the application of this approach was given in the New Zealand SPRFMO bottom fishing impact assessment (Ministry of Fisheries 2008), which included figures of the distribution of large seamounts at appropriate depths for certain commercial species, and the modelled habitat suitability of seamounts for stony corals.

The FAO Guidelines also provide definitions of species or communities that are considered to be vulnerable to impacts of bottom fishing, and examples of taxonomic groups of organisms which have those characteristics, and which could contribute to forming VMEs. The 4th meeting of the SPRFMO Scientific Working Group (SWG) noted that an effective approach to protecting such VMEs in the high seas SPRFMO Area:

> "would probably require the establishment of specific spatial closures designed to protect adequate and representative proportions of various VMEs. However, the SWG specifically emphasized that it would be critically important to ensure that any such spatial closures were based on best available information on distribution of VMEs, biodiversity and relative sensitivity of various habitat types to the impacts of fishing". (SPRFMO 2007)

Thus, SPRFMO has a clear requirement for robust and reliable models of the distribution of VME taxa in the South Pacific Ocean.

1.3 USA / New Zealand Joint Commission Meeting

New Zealand and the USA hold Joint Commission Meetings (JCM), approximately every two years, to discuss opportunities for collaboration under the USA-NZ bilateral Science & Technology Cooperation Agreement signed in 1974 and renewed in 1991. The JCM aims to:

- facilitate and promote enhanced RS&T cooperation between US and New Zealand at government, institutional and individual scientist levels; and

- contribute to and develop partnerships around common RS&T interests in cooperative activities over the next two-three years.

The 2nd JCM meeting, held in New Zealand in January 2010, incorporated an Ocean and Marine Sciences workshop to review latest developments and to discuss opportunities for collaborative research on identification, taxonomy and core data sets for deep-sea corals; identification, ecology and conservation science for vulnerable marine ecosystems; and habitat suitability and predictive modelling of deep-sea coral.

The following were among the key recommended areas for increased cooperation and activity reported from the workshop (Tracey et al. 2010):

- Mobilise existing unprocessed [VME] data sets and samples.
- Collate existing bathymetric/multibeam maps to help define VME likelihood.
- Share access to current 'best' environmental data layers; and encourage/contribute to development of new environmental data layers (e.g. bathymetry) and data sources (e.g. fishing industry).
- Update information on VME species distribution and association with environmental variables.
- Produce predictive models for VME taxa and use to assess effectiveness of VME management in the Pacific.
- Validate predictive models (field ground-truthing).
- Integrate modelling efforts predicting species distribution into fisheries management (including VME process in RFMOs) and conservation.

The JCM has identified collaborative research activities to collate environmental and VME taxa data, and to develop models to predict the distribution of VMEs with the ultimate aim of using such models in fishery management, as the key cooperative action item under the JCM Ocean and Marine Sciences Action Area in the JCM 'Roadmap of Cooperative Activities 2010 – 2020'.

1.4 The South Pacific VME research project

In order to address SPFRMO's specific requirements for robust and reliable models of the distribution of VME indicator taxa in the South Pacific Ocean, and to support the JCM's desire to promote collaborative research for the development of such models, a multi-lateral and multi-phased research project was proposed to develop predictive models for VMEs in the South Pacific Ocean, and to use these models, in conjunction with appropriate spatial management planning tools, to identify areas of the South Pacific that most warrant protection from the impacts of fishing activities.

It was proposed that the first phase of this project would focus on identifying, sourcing, collating and evaluating existing data sets that are likely to be useful for developing predictive models for VMEs across the South Pacific Ocean, and testing candidate predictive modelling approaches using these data. Future phases will obtain additional data, improve and validate preliminary models, and undertake analyses designed to inform discussions on the selection of spatial closures for the protection of VMEs in the SPRFMO area.

1.5 Objectives

The overall objective of Phase 1 of the project was to develop and test spatial habitat modelling approaches for predicting distribution patterns of vulnerable marine ecosystems in the Convention Area of the South Pacific Regional Fisheries Management Organisation with agreed international partners. There was one specific objective:

To collate data sets and evaluate modelling approaches which are likely to be useful to predict the distribution of vulnerable marine ecosystems in the South Pacific Ocean region.

2. METHODS

2.1 Vulnerable marine ecosystem indicator taxa

The FAO (2009) has provided guidelines and examples of taxonomic groups that could be used to indicate the presence of a VME, while noting that the list of taxa should be adapted to regional fisheries and ecologies. Parker et al. (2009) identified VME indicator taxa for the South Pacific Ocean based upon a number of ecological criteria: whether taxa were typically fragile relative to trawl gear; functionally significant to the community or ecosystem; unique, rare or endemic to a small area; or low productivity species due to life history traits such as slow growth, high longevity, low fecundity or unpredictable recruitment. In addition, they used two further pragmatic criteria: taxa needed to be retained to some degree and have been previously observed in bottom trawl gear and taxa must be readily identifiable by scientific observers on board without the aid of complex characters. These criteria identified eight vulnerable taxa. Parker et al. (2009) also selected two indicator or proxy groups, i.e. taxa typically found in association with VME indicator taxa. These ten taxa (Table 1) were proposed as the most suitable for use in the interim move-on rule implemented to prevent damage to VMEs in the South Pacific (Penney et al. 2009). These VME indicator taxa (Figure 1) were adopted here for use in this Phase 1 of the South Pacific VME Project.

2.2. Geographic scope

Given the known scarcity of data for much of the South Pacific Ocean region, and in order to make the running of preliminary models tractable, the first phase of the project focused on areas for which most data are available, and within which high seas fishing footprint mapping and VME protection measures have already been implemented by New Zealand. Thus the study area included the New Zealand region charting area (centred on the EEZ) and those regions of the high seas considered by Penny et al. (2009) (see their figure 1). This is an area bounded by coordinates: 24.00° S, 57.50° S, 157.00° E, and 160.00° W (Figure 2).

2.3 Data

2.3.1 VME indicator taxa data

Data for the ten VME indicator taxa were obtained from database sources in New Zealand and overseas. These were: *Specify*, *trawl*, *cod*, *AllseaBio*, and *marine_db* (national), and *OBIS*, and *Hexacorallians of the World* (international). Data were compiled, converted to a standard format, and duplicate records or errors removed. The final dataset contained over 31 000 individual records for Actinaria, Alyconacea, Antipatharia, Brisingida, Crinoidea, Gorgonacea, Pennatulacea, Porifera, Scleractinia, and Stylasteridae. Records were unequally distributed among the taxa (Table 2) and across the study area (Figure 3a-j), and data gaps (by area) were visually assessed by mapping the sampling stations in the study area from which records of the ten VME indicator taxa were obtained. The final dataset for the VME indicator taxa will be made available via a web portal on completion of the South Pacific VME Project.

2.3.2 Environmental data

Data for environmental variables were obtained from various sources. The variables were grouped into two environmental data sets: (1) a ‘regional’ set of environmental data layers that included variables developed or tuned specifically for the New Zealand region plus a limited number obtained from other mostly regional sources (e.g. CSIRO Atlas of Regional Seas); (2) a ‘global’ set of environmental data layers that represent a sub-set of data collated and processed by Davies & Guinotte (2011) from global data sources (e.g. World Ocean Atlas) (see Tables 3 and 4 for full details

of data sources). All these variables were selected based on the knowledge or assumption that they directly or indirectly influence the distribution of VME indicator taxa across the study area.
Eleven environmental variables were used in the regional dataset: depth, bottom water temperature, slope, seamount, current speed at the seafloor, tidal current speed, dynamic topography, sea surface temperature gradient, surface water primary productivity, dissolved organic matter, and particulate organic carbon flux to the seafloor.

The depth layer used was interpolated from multi-beam and single-beam echo sounders, and satellite gravimetric inversion, at a 1 km^2 grid-scale resolution. Although depth is not a direct driver of species distributions, it was included as a surrogate for other variables that are known to influence species distributions in the deep sea, e.g., pressure, light, and oxygen (Levin et al. 2001). Temperature is known to influence the physiology and thus the distribution of deep-sea species (Thistle 2003). Seabed slope, a proxy for a range of potentially suitable environmental conditions for fauna that inhabit steep or gently sloping seafloor, was determined from the bathymetry layer using neighbourhood analysis (determined at the 1 km scale). For example, substratum type, which is one of the most important variables controlling the distribution of seafloor organisms (Snelgrove & Butman 1994), is likely to be fine mud or sand where the slope is flat or gentle and sediment can remain deposited. Where the slope is steep, bottom current speeds can be accelerated, sediment is less likely to settle, and the seafloor is more often bare rock or composed of coarser substrates. Another habitat layer, ‘seamount’, was used to describe seamount occurrences in the study area; the summits and upper slopes of seamounts are known to be habitats suitable for VME indicator taxa such as corals (Genin et al. 1989).

Measures of potential food availability were described indirectly by four variables, assuming that surface water primary productivity will influence deep-water productivity due to the settling of detritus to the seabed which could provide a food source for VME indicator taxa such as corals and sponges (Duineveld et al. 2004, Duineveld et al. 2007). These variables included: sea surface temperature gradient, which indicates the major zones of water mixing (fronts) that are often associated with increased productivity; surface water primary productivity, derived from a vertically generalized productivity model based on remotely-sensed estimates of chlorophyll a concentration; dissolved organic carbon, also estimated from ocean colour data with an atmospheric correction and optical property algorithm; and particulate organic carbon flux, described as a function of the production of organic carbon in surface waters or the export of organic carbon from the base of the euphotic zone scaled to depth below the sea surface. Water current regime is thought to be important to benthic organisms, principally as a delivery mechanism for food material (Genin et al. 1989, Thiem et al. 2006), but can also reflect the degree of physical disturbance to which seafloor animals are subjected and, as a consequence, their distribution (Warwick & Uncles 1980, Long 1997), as well as represent a proxy measure of substratum type.

Current variables used in the analysis included dynamic topography, current speed at the seafloor, and tidal current speed. The mean dynamic topography variable is related to mean surface current by the geostrophic relationship, i.e., current flows along the isolines of the dynamic topography. Surface currents may be responsible for moving potential food material from one location to another, whilst bottom currents may have a more direct influence upon the availability of food for benthic organisms at the seabed. All these data (apart from seamount and bathymetry data) represent a ‘climatology’, i.e. they are average values over a period that is effectively determined by when the observations were made, primarily the last couple of decades. All regional environmental variables were mapped in a World Mercator projection (central meridian 100°E, standard parallel 41°S) at 1 km^2 resolution (Figure 4).

Nine environmental variables were used in the global dataset: depth, bottom water temperature, particulate organic carbon flux to the seafloor, slope (determined at 5 km scale), bottom water salinity, dissolved oxygen concentration, calcite saturation state, aragonite saturation state, and silicate. These variables were included because they were strong predictors in the global distribution of deep sea coral and sponge ecosystems (Davies & Guinotte 2011). This was especially true for aragonite saturation state, dissolved oxygen, slope, and silicate. The choice of variables used for each taxa model varied according to the known and/or suspected controls on their distribution. For example, carbonate chemistry variables were used for deep-sea coral taxa predictions as they exert a strong control on their distribution, but they are not known to control the distribution of sponges. Silicate was used in sponge prediction models as sponges need silicate to build their internal structures.

2.4 Predictive modelling approaches

A number of methods have been used recently to model the distribution of a variety of marine taxa in environmental space (i.e., to determine habitat suitability). Examples for coldwater corals alone include; Bryan & Metaxas (2006), Bryan & Metaxas (2007), Davies et al. (2008), Tittensor et al. (2009), Woodby et al. (2009), Tittensor et al. (2010), Davies & Guinotte (2011), Tracey et al. (2011a), Yesson et al. (2012). However, the ability of such models to act as good predictors of biological patterns is strongly dependent on how they are constructed (e.g. the modelling technique) and the environmental data used (Araújo & Guisan 2006, Guisan et al. 2006). The suitability of two modelling approaches was evaluated for predicting the distribution of VME indicator taxa in the South Pacific region.

2.4.1 Maximum Entropy Modelling

One of the most popular habitat suitability modelling methods is maximum entropy modelling (Maxent) (e.g., Dolan et al. 2008, Guinan et al. 2009). Maxent is a presence-only approach that has been shown to out-perform some other presence-only techniques including Ecological Niche Factor Analysis (ENFA) (e.g., Tittensor et al. 2009). The underlying assumption of Maxent is that the best approach to determining an unknown probability distribution is to find the solution which has maximum entropy based on constraints derived from environmental variables (Phillips et al. 2006). The algorithm is supplied within a Java software package (Maxent version 3.2.1), and the default model parameters of Maxent have been found to perform well in other studies (Phillips & Dudik 2008). Covariation between environmental data is a complication that must be addressed in many predictive modelling efforts. Environmental data used in Maxent analysis can be assessed for covariation in a correlation matrix. Strong correlations between variables (correlation over 0.7) can be addressed by omitting one of the environmental variables. The importance of each variable in the model is assessed using a jack-knifing procedure that compares the contribution of each variable (when absent from the model) with a second model that includes the variable. The final habitat suitability maps are produced by applying the calculated models to all cells in the study area, using a logistic link function to yield a habitat suitability probability between zero and one (Phillips & Dudik 2008).

Several studies have highlighted issues with using only one statistic to evaluate model performance (e.g. Elith & Graham 2009). Model accuracy between the test data and the predicted suitability models can be assessed using a threshold-independent procedure that uses the area under the receiver operating characteristic curve (AUC) for the test localities and a threshold-dependent procedure that assesses misclassification rate. To calculate validation metrics, the presence data are randomly partitioned to create 75% training and 25% test datasets, with test data used to calculate validation metrics. With presence-only data, Phillips et al. (2006) define the AUC statistic as the probability that a presence site is ranked above a random background site; AUC scores of 0.5 indicate that the discrimination of the model is no better than random, and the maximum AUC value is 1, indicating highest probability of suitable habitat (Hanley & McNeil 1982). Several studies have criticised the use of AUC as a single metric for assessing performance because AUC is sensitive to the total spatial extent of the model (Lobo et al. 2008, Petersen et al. 2008). Two further metrics can be applied when using Maxent, (1) a threshold-dependent omission rate (fixed value of 10) (Pearson et al. 2007), which evaluates model success by assessing the proportion of test locations that fall into cells that were not predicted as suitable, and (2) Test gain, which can be interpreted as the average log probability of the presence samples used to test the model. For example, if the test gain is 2, the average likelihood of a test presence locality is e^2 (about 7.4) times greater than that of a random background pixel (Riordan & Rundel 2009).

2.4.2 Boosted Regression Trees

Another recently developed and adopted approach for modelling taxa distribution is boosted regression trees analysis (BRT) (De'ath 2007, Elith et al. 2008). Boosted regression trees, an advanced form of regression analysis, have shown a higher performance and provide a number of advantages over standard regression methods, such as generalized additive models (Leathwick et al. 2006, Elith et al. 2008). Regression trees can model almost any type of predictor variable, can automatically model interactions between variables, and can deal with differing scales of measurement among predictors. Boosting improves the performance of standard regression trees by stochastically fitting the data such that the emphasis during model building is placed on explaining the unexplained variation in the data space at each step (Elith et al. 2008).

The influence of the environmental variables to the explanatory power of each BRT model can be assessed using functions provided by Elith et al. (2006). Model performance can be assessed using cross-validated measures of deviance, discrimination, and correlation. Model performance can also be assessed during cross-validation by comparing model predictions to withheld portions of the data (Elith et al. 2006). The null deviance provides a measure of variation where there are no covariates, i.e. a model with a constant term of 1, and the explained deviance gives a measure of the goodness-of-fit between the predicted and raw values. The percentage of the deviance explained is the deviance expressed as a percentage of the null deviance for each species (Leathwick et al. 2008). The degree to which fitted values discriminate presences and absences in the model is measured by the AUC performance metric. To identify the final set of variables that optimise the model, a model simplification code that sequentially drops variables that do not change the cross-validated deviance of each model can be used (Elith et al. 2008).

2.4.3 Target-group background or pseudo-absence data

Ideally, predictive models need a measure of sample absence, but in cases where such data do not widely exist (e.g. the deep sea) pseudo-absences can be used instead to provide a measure of the background environment across the sampled region (Phillips et al. 2009). Target-group background or pseudo-absence data can be provided by location records from research surveys that were carried out over the same period as the taxa data collection, but where no taxa were recorded when using the same sampling gear types. Because these targeted pseudo-absences reflect the background environment where taxa are not observed, their inclusion can enable the model to focus on distributional differences between the recorded observations and the background data, and not include sample selection biases (Phillips et al. 2009). If the number of pseudo-absences outnumbers the number of presences in the model for each taxon, then the overall influence of the pseudo-absences can be down-weighted such that their weights are constant, and scaled to equal the sum of the presence observations. The alternative to the use of target-group background data is to simply use a number of randomly selected points in the modelling domain to generate background data.

2.5 Model Analysis

Three types of model were made for each VME indicator taxon: (1) a Maxent and (2) a BRT model using the ‘regional’ environmental data, and (3) a Maxent model using the ‘global’ environmental data.

The first two model types allow for a performance evaluation of the two different modelling approaches, while comparing the first model type with the third model type allows for an assessment of the importance of including regionally-tuned environmental layers in models of VME indicator taxa (even though the environmental variables are not strictly matched). The performance evaluations were made using performance measures specific to each modelling approach and the AUC measure, a quantitative performance measure common to all model types. The model map outputs were also compared qualitatively to assess the potential influence of using different models, or different environmental data, on the predicted distribution of suitable habitat for VME indicator taxa in the study area.

Finally, to assess the usefulness of producing models of VME indicator taxa at the taxonomic level identified by Parker et al. (2009), the predicted distribution of habitat suitability for Scleractinia was compared to that of a single species of habitat-forming stony coral, *Solenosmilia variabilis* (data from Davies & Guinotte 2011). Both of these models were made using Maxent with global environmental data.

2.5.1 Maxent analysis

Maxent version 3.2.1 was used to predict the potential distribution of VME indicator taxa using both ‘regional’ and ‘global’ environmental data. Default model parameters were used as they have performed well in other studies (a convergent threshold of 10-5, maximum iteration value of 500 and a regularization multiplier of 1 (Phillips & Dudik 2008, Davies & Guinotte 2011)). Target-group background data were not used for the Maxent analysis, instead background data used for each model analysis was comprised of 10 000 randomly selected points throughout the maximum extent of the study area. The selection of the number of random points is a user driven process and n=10 000 has performed well in other modelling efforts (Davies & Guinotte 2011, Guinotte & Davies 2012, Yesson et al 2012). Duplicate taxa records (multiple records occurring in the same geographic cell) were removed prior to modelling. Duplicates were removed to prevent sampling bias from skewing model results towards environmental conditions found in heavily sampled areas. To calculate validation metrics, the presence data for each VME indicator taxon were randomly partitioned to create 75% training and 25% test datasets, with test data used to calculate performance metrics. Model accuracy between the test data and the predicted suitability models was assessed using the threshold-independent AUC procedure, test gain, and a threshold-dependent procedure that assessed misclassification rate.

Maxent model results (probabilities from 0–1) for each taxon were mapped using ArcMap 10 GIS software (www.esri.com). Null values were displayed as white grid cells.

2.5.2 BRT analysis

BRT analyses were run in R (R Development Core Team 2012) using the library *gbm* and functions described by Ridgeway (2006), Elith et al. (2008), Leathwick et al. (2008), Elith & Leathwick (2011). The BRT models used a binomial error distribution (family Bernoulli) to predict the habitat suitability or probability of occurrence of each of the VME indicator taxa. Two of the primary factors that control the BRT model fit – the ‘learning rate’ and the number of trees – were optimised within the model. The third factor, the number of interactions that determine a split (“tree complexity”) was set to 2, where 1 means that there are no interactions. Allowing interactions reflects the understanding that at least some of the environmental variables in combination may create an environment that influences the response variable (the taxon). The effect of each the environmental predictor variables in the model is presented as their relative contribution (%). Two cross-validation measures were estimated within the model (using 75% training, 25% test data): the percentage deviance explained, and the AUC value.

For the BRT analysis, it was assumed that the VME indicator taxa represented the full sample identified from each station and that a null record for a taxonomic group was an absence for that taxon. Target-group background data, or pseudo-absence data, were provided by an expanded version (about 55 000 stations) the “all benthic stations” dataset of Tracey et al. (2011a). The target-group dataset details the location of all stations in the study area that have been sampled with gear types (sleds, trawls, grabs etc) that provided records for the VME indicator taxa records, but at which no VME indicator taxa were recorded.

All records in the VME indicator taxa and target-group background datasets were assigned to the grid of 1 km^2 cells that represented the study area. For both the presence and background data, all records in each cell were used in the BRT analysis (in contrast to the Maxent approach). By incorporating all the records, the density of presences and target-group background data provided more information about the environmental space (habitat) preferred by each taxon.

For the prediction, the environmental variables were combined into one dataset that represented the study region, with each cell represented by location information. Any cells with null values were deleted from the dataset. When displayed in GIS, these excluded cells were restricted to shallow waters close to land. The predict.gbm function in R was used to generate a dataset of the probability of habitat suitability for each VME taxa, using the environmental variable dataset. These prediction datasets were exported and displayed in GIS for interpretation. Maps and GIS analyses were done in ArcMap 10 GIS software (www.esri.com).

3. RESULTS

Appendix 1 contains figures that show the distribution of habitat suitability for all VME indicator taxa within the study area, as predicted by the three types of model. Tables in Appendix 2 list the environmental variables that contribute the most to the models that explain the distribution of the habitat suitability for the VME indicator taxa.

Below, a selection of figures from Appendix 1 are used to illustrate the results of the model performance evaluation, assessment of the importance of the ‘regional’ versus ‘global’ environmental layers for model performance, and the usefulness of using a high taxonomic level for VME indicator taxa. No further reference is made to the variables that explain the different models, as these results are irrelevant to the objectives of the present study.

3.1 Maxent and BRT model performance

The performance measures for the Maxent models indicate that these models all performed well. Omission rates were low (no more than 3.2%), Test gain was relatively high (over 2.0), as were the AUC values (over 0.96) (Table 5). These results indicate that only a few presences were misclassified as absent and that predicted presences were orders of magnitude more probable than that of a random background pixel. The performance measures for the BRT models also showed that these models generally performed well. The % deviance explained was mostly more than 0.3, and the AUC values were all more than 0.7 (the cut-off for “useful” models, Swets 1986) (Table 6). These results indicate that, in general, the BRT models predicted the subsets of withheld data well. However, model performance did vary noticeably by taxa (Table 6). The BRT model for Antipatharia was the best performing model (% deviance explained, 0.51; AUC, 0.91), and the model for Alcyonacea was the worst performing model (% deviance explained, 0.23; AUC, 0.77).

The distribution maps of predicted habitat suitability produced by the Maxent and BRT models for the ten VME indicator taxa were almost all qualitatively dissimilar to one another (e.g. Crinoidea, Figure 5a), with only the map for Pennatulacea showing a somewhat similar distribution of high and low habitat suitability throughout the study area (Figure 5b). In general the Maxent predictions of suitable habitat are constrained to areas where data records exist, while BRT predictions of suitable habitat extend more into unsampled areas (e.g. Figure 5). BRT models also produce greater variability in their predictions of habitat suitability at the smallest spatial scales than Maxent models (i.e. patchiness) (e.g. see Scleractinia, Figure 6).

3.2 Regional and global environmental data

A comparison of the AUC metric for the Maxent models using ‘regional’ and ‘global’ environmental data layers indicated that the two models perform equally well. With the exception of one model (Actinaria, 0.88; based on ‘global’ environmental data), all models had AUC scores of at least 0.95 (Table 7).

The distribution maps of predicted habitat suitability produced by the Maxent models for the ten VME indicator taxa based on the ‘regional’ and ‘global’ environmental data layers were overall qualitatively similar (e.g. Brisingida, Figure 7a). However, for two taxa (Pennatulacea and Stylasteridae) the models based on the ‘regional’ environmental data layers predicted areas of high habitat suitability (over 0.8 probability) not predicted by corresponding models based on the ‘global’ environmental data (e.g. Pennatulacea, Figure 7b).

3.3 General usefulness of VME indicator taxa approach

All the BRT and Maxent models predicted that large areas of the study area represented suitable habitat for VME indicator taxa. For example, Maxent models using ‘regional’ and ‘global’ environmental data predicted that almost the entire study area at depths less than 750 m was suitable habitat (over 0.6 probability) for Actinaria, Porifera, and Scleractinia (e.g. Porifera, Figure 8). While the BRT models did produce predictions that discriminate relatively small areas of high habitat suitability (over 0.8 probability), only a few Maxent models for VME indicator taxa predicted discrete areas of high habitat suitability (see also above). The general inability of models to finely discriminate highly suitable habitat is in part a consequence of the coarse taxonomic levels used to classify VME indicator taxa. The models defined the environmental niche for a large number of species within a phylum (Porifera), class (Criniodea), order (Actinaria, Antipatharia, Brisingida, Gorgonacea, Pennatulacea, Scleractinia) and family (Stylastrida). When the species within these taxonomic groups have widely different habitat requirements, the models are unlikely to produce predictions that resolve areas of highly suitable habitat in fine detail. The comparison of the Maxent model prediction for all Scleractinia compared to *Solenosmilia variabilis* illustrated the issue of using coarse taxonomic levels for VME indicator taxa. The model for habitat-forming *S. variabilis* (Figure 9a) predicts discrete areas of highly suitable habitat within a broader area of suitable habitat, which is similar in extent to that predicted for Scleractinia as a taxonomic group (Figure 9b).

3.4 Data gaps

The total number of collated records varied considerably among the different VME indicator taxa, and the number of records that could be used for modelling purposes ranged from 505 to 9187 per taxon. The difference in the number of records to some extent reflects the level of taxonomic grouping (i.e., the largest number of records is for the phylum Porifera), but could also reflect the relative ubiquity of the group in the study area or their catchability by the commonly used sampling gears. However, there were sufficient numbers of records to produce useful models for all taxa (e.g. 20 records have been suggested as a minimum for Maxent models, Papeş & Gaubert 2007).

While the number of data records is apparently sufficient to produce useful models, the distribution of data records across environmental space can reflect a sampling bias that can impact upon model performance and spatial uncertainty. The sampling effort used to identify the presence and absence datasets is from over 50 years of data collection and reflects different surveying objectives: for example, certain seamount areas have been routinely and relatively intensively surveyed compared with other areas that may have been sampled on only one occasion. While an attempt was made to reduce the effect of sample bias on the BRT models (e.g. by using target-group background data, see above), the potential impact of sampling bias on the Maxent models was demonstrated by the predictions of habitat suitability being general constrained to areas for which there were data records. Figure 10 shows that VME indicator taxa data records used in the modelling were obtained from benthic sampling stations mainly concentrated in the New Zealand EEZ, with very few sampling stations located in the high seas portion (i.e. the SPRFMO area) of the study area. This map illustrates that there are some obvious gaps in the distribution of sampling stations from which VME indicator taxa records have been drawn. These areas include the Louisville Seamount Chain, West Norfolk Ridge, and Lord Howe Rise.

4. DISCUSSION

The first phase of the South Pacific VME Project successfully collated over 30 000 records for VME indicator taxa, obtained and developed data for a number of environmental variables, and evaluated two habitat suitability modelling approaches for predicting the distribution of ten VME indicator taxa in the South Pacific Ocean region.

The study provided answers to a series of questions that inform the next planned phases of the overall project.

4.1 Which modelling approach is best?

Both types of modelling approach produced well-performing models, a result consistent with direct comparisons of performance that included BRT and Maxent models (Elith et al. 2006). However, the predicted distribution of habitat suitability for VME indicator taxa in the study area did differ noticeably between the two model types. This dissimilarity relates to the difference in the operational choices and statistical methods which underlie the two approaches used.

The BRT approach included the use of target-group background (or pseudo-absence) data drawn from the positions of benthic sampling stations, whereas the Maxent approach adopted used randomly selected background data locations from the study area. The Maxent procedure used only one presence record for a VME indicator taxon per grid 1 km^2 grid cell, whereas the BRT models used all data records within a cell for a particular taxon. Thus, the BRT model input data were more informative about preferred space and by using pseudo-absence data there is additional information for each location and the prevalence of taxa (Phillips et al. 2009). Together these differences in the modelling approach are responsible for producing the dissimilarity in the predicted distributions of habitat suitability in the study area. These included: the finer scale spatial variability observed for the BRT models; and the constraining of the Maxent models to areas where the majority of data records were located, while BRT predicted into the unsampled deeper regions of the study area.

The latter observation has implications for the use of these two modelling approaches for future phases of the project, i.e. predicting suitable habitat for VME indicator taxa in the entire SPRFMO area. Most of the SPRFMO region is unsampled (see Section 4.4) and thus Maxent may not be the best choice of modelling approach if it does not generally predict suitable habitat for VME indicator taxa beyond the sampled domain. However, while the BRT models did predict suitable habitat outside of the sampled domain, several of the predictions of suitable habitat in the deeper unsampled regions appeared to be implausible.

BRT and Maxent are widely considered in the ecological modelling literature as among the most robust species distribution/habitat suitability modelling approaches currently available (Elith et al. 2006, Phillips & Dudik 2008, Elith & Graham 2009). Maxent's use of presence-only data has already led to its wide acceptance for conservation and management applications (see examples in Elith et al. 2011). BRT models using presence-only data are increasingly being incorporated in ecological studies (e.g. Compton et al. 2013), and have also begun to be used recently to model the distribution of species of conservation significance (e.g. Torres et al. 2013). The use of presence-only data is the norm for habitat suitability modelling of the under-sampled deep sea (e.g. Bryan & Mataxas 2007, Woodby et al. 2009, Tittensor et al. 2009, Yesson et al. 2012), and for situations that require a useable model based on such data in the face of a pressing management need (e.g. Clark & Tittensor 2010, Beaugrand et al. 2011).

Habitat suitability modelling approaches are rapidly developing and there is a growing number of different model types that produce high performing models with typical marine data, and which could be suitable for modelling VME indicator taxa, or combinations of taxa (e.g. Dunstan et al. 2011ab, Pitcher et al. 2012). Given the availability of new modelling approaches, and the concerns raised above about the outputs generated by Maxent and BRT in this study, it would be prudent for the next phase of the South Pacific VME project to continue to evaluate the suitability of different methods for the types of data available and the predictive outputs required.

4.2 Can ‘global’ environmental data layers be used?

The Maxent models constructed using ‘regional’ and ‘global’ environmental data layers preformed equally well, and produced broadly similar predictions of the distribution of habitat suitability for the VME indicator taxa across the study area. That said, two of the ten models made using ‘regional’ environmental data did predict some discrete areas of high habitat suitability that were not predicted by the corresponding models based on ‘global’ data. This relatively small difference in the predictions for these two taxa could relate to the use of the regionally-tuned data layers, but it could also be a consequence of the type of comparison made. The comparison between the two types of Maxent model was not made using exactly the same set of environmental variables. While both sets of environmental data layers contained some of the same variables (depth, temperature, slope) derived from different sources (‘regional’ versus ‘global’), they also each contained an independent suite of a different number of other variables. So it is possible that the difference in the types and number of variables used could account for the ability of the Maxent model based on ‘regional’ data to discern high habitat suitability for two VME indicator taxa.

That the two sets of environmental data layers produced similarly-performing models was somewhat contrary to what was expected, in that it would be reasonable to predict that the use of some regionally-tuned environmental data layers would noticeably improve model performance. So despite the partially confounded nature of the comparison undertaken, it served its purpose in revealing the usefulness of using ‘global’ environmental data layers in habitat suitability models of restricted regions of the world’s oceans. Because the models based on the ‘global’ environmental data had very high AUC performance scores and maps of habitat suitability were similar overall to those produced using ‘regional’ data, it is reasonable to conclude that future habitat suitability models for VME indicator taxa in the South Pacific can make good use of environmental data that is already available, rather than spending valuable resource on developing new regionally-tuned environmental data layers.

The future phases of the South Pacific VME Project can use a set of 30 global environmental data layers collated and developed by Davies & Guinotte (2011), and possibly other data being collated for the Bio-ORACLE global marine dataset (Tyberghein et al. 2011). However, there is at least one valuable data layer that is still lacking at large spatial scales. While the environmental data layer ‘slope’ is a useful proxy for seafloor substratum type, the lack of a robust global- or even regional-scale data layer for substratum type is lamentable considering the influence of this variable on the distribution of seafloor organisms. The modelling of habitat suitability for bryozoans in the New Zealand EEZ, using far from perfect substratum type data layers derived from spatially inconsistent sediment charts (Leathwick et al. 2009), has recently demonstrated the importance of including substratum variables to improve model performance (Wood et al. 2013). Data for seafloor sediments are available in various regional/global databases (e.g. dbSEABED, http://instaar.colorado.edu/~jenkinsc/dbseabed), and it is possible to use these data to make spatially consistent and continuous interpolated maps or models of substratum characteristics (e.g. mud content, Li et al. 2011). Thus, it is important to develop an environmental data layer for substratum type for future modelling of VME indicator taxa in the South Pacific.

4.3 Is the general modelling approach sensible?

The modelling evaluation demonstrated the lack of fine spatial resolution in predictions that result from using data for high taxonomic levels in habitat suitability modelling. Using taxonomic data collated at the phyla (e.g. Porifera) or order (e.g. Actinaria) level is bound by the 'environmental niche' nature of the BRT and Maxent modelling approaches to produce broad predictions of habitat suitability. So while the current taxonomic levels for VME indicator taxa in the South Pacific (Parker et al. 2009) are appropriate for implementing the protection measure for which the criteria were designed ('move-on' rule based on observer identifications of VME indicator taxa), these levels are inappropriate for predicting the likely occurrence of VMEs in the SPRFMO region to design spatial management measures such as area closures. The latter will require high spatial resolution predictions of suitable habitat for VME indicator taxa, most likely at the scale of individual seamounts or groups of seamounts (kilometres to tens of kilometres). The present study demonstrated, using models for Scleractinia and *Solensmilia variablis,* that habitat suitability models for species will produce predictions at these scales. Thus, before the future phases of the South Pacific VME Project begin it will be necessary to revise criteria for identifying suitable VME indicator taxa. For example, it may be necessary to develop criteria that identify key habitat-forming taxa at the species level, or at the genera level when a genus is a useful proxy for a group of species that are indicative of a VME, while having similar environmental requirements for their occurrence. Clearly it will first be necessary to determine whether sufficient species and genera level data are available across the SPRFMO region, otherwise modelling VME taxa will be possible at high taxonomic levels only.

The measurement or modelled scale of the original data used to generate the environmental data layers used in models will also affect the resolution of the predicted distribution of habitat suitability (Araújo & Guisan 2006). Unfortunately, very little environmental data that describes the seafloor or ocean environment is originally collected at scales of kilometres or even tens of kilometres across large areas, and so there is an intrinsic spatial resolution limit to regional models based on the majority of available environmental data layers for the deep sea. The exceptions are bathymetric and backscatter data that can be collected by multibeam surveys at scales of 25 m^2. Terrain (e.g. slope, rugosity, aspect etc) and substrate (e.g. relative seafloor 'reflectivity') metrics derived from such surveys have been used successfully for habitat modelling of relatively small areas (e.g. Dolan et al. 2008), and more recently as multibeam surveys are conducted for large areas of shelf and slope, for larger areas. For example, recent work in the northeast Atlantic has demonstrated the usefulness of terrain and substrate metrics derived from multibeam surveys to produce high resolution habitat suitability models of VME indicator taxa (Howell et al. 2011, Rengstorf et al. 2012). Multibeam surveys have been conducted by the fishing industry of prospective fishing grounds in the SPRFMO region, and such a practice is likely to continue as new areas of the region are explored for new fishing opportunities. Data from these multibeam surveys could be used to produce high resolution predictions of habitat suitability for VME indicator taxa, and in the future phases of the South Pacific VME project such models should be made for wherever there are appropriate multibeam data.

Howell et al. (2011) also investigated whether the predicted distribution of a habitat-forming coral species, *Lophelia pertusa*, acts as a suitable proxy for the reef habitat itself – the actual VME. Their study found that while highly suitable habitat for the species was predicted to occupy 7.17% of the study area, only 0.56% of the area was predicted to be suitable for the reef. These authors concluded that the difference in the distributions of reef and species suggest that predictive modelling efforts should focus on the habitat rather than the species at fine spatial scales (i.e. where there are multibeam data). Modelling the VME habitat itself (be it a deepwater coral reef or a sponge garden) in comparison to a single indicator species is eminently sensible where there are suitable data derived from video and multibeam surveys. However, data for VME habitats across areas as large as the New Zealand EEZ or the SPRFMO region are incredibly sparse (which of course necessitates the South Pacific VME Project) and so other approaches to more accurately predict the distribution of VMEs themselves should be considered in the future. For example, models which combine the predictions of multiple VME indicator taxa could improve the likelihood of identifying VMEs, particularly if they used data records collated at fine taxonomic levels. The modelling of multiple taxa to predict the location of VMEs also presents the opportunity to use modelling approaches that have recently been used with some success to map the distribution of seafloor assemblages or communities (e.g. Generalised Dissimilarity Modelling, Compton et al. 2013; Gradient Forest, Pitcher et al. 2012; finite mixing model, Dunstan et al. 2011a). Such modelling approaches (and BRT) can also use abundance data and provide measures of model uncertainty which can be spatially represented on a map. The ability to use abundance data is an attractive proposition, for the presence of a VME is often related to the relative abundance of key habitat-forming species (e.g. coral colonies). However, abundance data are relatively sparse for VME indicator taxa that were gathered for this study, and such records may only be reliably recovered from video surveys. Producing predictions for habitat suitability with measures of uncertainty is also an ideal, because decision-making around the selection of protection areas should be made with an understanding of the uncertainty attached to the likelihood of a VME occurring in a particular area.

Recent research using species distribution modelling is beginning to exploit the growing amount of information that is available on the genetic structure of populations (e.g., Hamann & Aitken 2013). These approaches allow for the inclusion of data on genetic connectivity between spatially distinct populations. Incorporating connectivity information into predictions of habitat suitability for VME indicator taxa would provide useful information about the genetic uniqueness of a particular population as well as the likely recovery of any VMEs disturbed by fishing. It is easy to envisage how such information could be used in the spatial management planning process to identify suitable areas for protection.

4.4 Where are new data needed from?

Data for VME indicator taxa are generally sparse outside of the New Zealand EEZ, and particularly sparse in those areas where the New Zealand fishery is concentrated in the SPRFMO region: the Louisville Seamount Chain; West Norfolk Ridge; and Lord Howe Rise. These data gaps may impact upon the performance of future habitat suitability models for areas (and related depths) of the SPRFMO region that are of particular significance to the fishery. While new data will be forthcoming from the EEZ for the next phases of the South Pacific VME Project from planned biodiversity and fisheries bycatch surveys, the availability of data from outside of the EEZ will be much more limited. One important source of data for VME indicator taxa from high seas will be those data collected by the fisheries observer programme. However, the usefulness of these bycatch data may be restricted by the error and taxonomic level restriction associated with observer-based identifications, and the relatively small amount of such identifications that are subsequently subjected to verification by taxonomic experts (Tracey et al. 2011b). Thus, in order to obtain good data for VME indicator taxa from areas that require the most robust habitat suitability predictions, it is desirable to undertake a dedicated biodiversity survey to areas such as the Louisville Seamount Chain. This survey should also collect multibeam data if it has not been collected previously by the fishing industry. Such a survey would also serve to ground-truth preliminary models and develop future models for VMEs with the greatest degree of reliability.

The survey offers the opportunity for further US-New Zealand collaboration under the JCM agreement to improve coral data and identify VMEs using habitat suitability models in the Pacific Ocean. This survey could also promote collaboration with nation states in the region that are party to the SPRFMO agreement (e.g. Australia).

5. RECOMMENDATIONS

The results of the Phase 1 of the South Pacific VME Project, and the above discussion, provide recommendations that can inform the future phases of the project:

- Take both BRT and Maxent modelling approaches, and others, forward to the next phase of project.
- Revise the criteria for identifying VME indicator taxa, e.g. species level for key habitat-forming taxa.
- Generate predictive models that combine the presence of particular taxa to help refine the identification of VMEs.
- Where multibeam surveys have been undertaken in the SPRFMO region, bathymetric and backscatter data should be used to make high resolution habitat suitability maps.
- Undertake a biodiversity survey of the Louisville Seamount Chain (and/or West Norfolk Ridge, Lord Howe Rise) to obtain data for VME indicator taxa and habitat, ground truth the preliminary models, and develop refined models.
- New models for the SPRFMO region should include those that model VME habitat (e.g. deepwater coral reef) directly, and if possible incorporate estimates of genetic connectivity, as well as spatially explicit measures of uncertainty.

6. ACKNOWLEDGEMENTS

Thanks are owed to Brian Sanders, Brent Wood and Kareen Schnabel (NIWA) for assistance with biological data extracts. Tanya Compton (NIWA) is acknowledged for carrying out the initial modelling that was ultimately replaced by models produced from extended regional environmental data layers. The use of these extended regional environmental layers was not originally envisaged, but when the opportunity to include bathymetric data obtained by the fishing industry from the Louisville Seamount Chain (supplied via Andrew Penney, MPI) arose, the study area was expanded and regional environmental data layers were further developed. Mark Hadfield, Mark Gorman, and Matt Pinkerton are thanked for their skill and efforts in the development and production of these data layers. We would also like to thank Andrew Davies (Bangor University) for his technical expertise and analytical work to produce and improve the global environmental database that made this work possible. Funding for this project was provided by the MPI (project number ZBD2010-40), which was augmented by NIWA so that data from the fishing industry data could be incorporated into the bathymetric data layer, and other environmental data layers could be expanded in extent. Financial support (for modelling carried out by John Guinotte, Marine Conservation Institute) was also provided by NOAA Fisheries Office of International Affairs, NOAA's Deep Sea Coral Research and Technology Program, and the National Fish and Wildlife Foundation We are grateful to Andrew Penney and Mary Livingston (MPI), and Malcolm Clark and Sophie Mormede (NIWA) for useful discussions and support during the project. David Bowden and Barbara Hayden provided comments that helped to improve the first draft of the report.

7. REFERENCES

Araújo, M.B.; Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33:1677–1688.

Auster, P.J.; Gjerde, K.; Heupel, E.; Watling, L.; Grehan, A.; Rogers, A.D. (2011). Definition and detection of vulnerable marine ecosystems on the high seas: problems with the “move-on” rule. *ICES Journal of Marine Science*, 68(2), 254–264.

Beaugrand, G.; Lenoir, S.; Ibañez, F.; Manté, C. (2011). A new model to assess the probability of occurrence of a species, based on presence-only data. *Marine Ecology Progress Series* 424: 175–190.

Becker, J.J.; Sandwell, D.T.; Smith, W.H.F.; Braud, J.; Binder, B. et al. (2009). Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30_PLUS. *Marine Geodesy* 32: 355–371.

Behrenfeld, M.J.; Falkowski, P.G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography* 42:1–20.

Boyer, T.P.; Levitus, S.; Garcia, H.E.; Locamini, R.A.; Stephens, C. et al. (2005). Objective analyses of annual, seasonal, and monthly temperature and salinity for the World Ocean on a 0.25u grid. *International Journal of Climatology* 25: 931–945.

Bryan, T.L.; Metaxas, A. (2006). Distribution of deep-water corals along the North American continental margins: Relationships with environmental factors. *Deep-Sea Research* I 53:865–1879.

Bryan, T.L.; Metaxas, A. (2007). Predicting suitable habitat for deep-water gorgonian corals on the Atlantic and Pacific Continental Margins of North America. *Marine Ecology Progress Series* 330:113–126.

Clark, M.R.; Tittensor, D.P. (2010). An index to assess the risk to stony corals from bottom trawling on seamounts. *Marine Ecology* 31(Suppl. 1): 200–211.

Compton, T.J.; Bowden, D.A.; Pitcher, R.; Hewitt, J.E.; Ellis, N. (2013). Biophysical patterns in benthic assemblage composition across contrasting continental margins off New Zealand. *Journal of Biogeography* 40(1): 75–98.

Davies, A.J.; Wisshak, M.; Orr, J.C.; Roberts, J.M. (2008). Predicting suitable habitat for the cold-water coral *Lophelia pertusa* (Scleractinia). *Deep-Sea Research* I, 55:1048–1062.

Davies, A.J.; Guinotte, J.M. (2011). Global habitat suitability for framework-forming cold-water corals. *PLoS ONE* 6: e18483.

De'ath, G. (2007). Boosted trees for ecological modeling and prediction. *Ecology* 88: 243–251.

Dolan, M.F.J.; Grehan, A.J.; Guinan, J.C.; Brown, C. (2008). Modelling the local distribution of cold-water corals in relation to bathymetric variables: Adding spatial context to deep-sea video data. *Deep-Sea Research* I 55:1564–1579.

Duineveld, G.C.A.; Lavaleye, M.S.S.; Berghuis, E.M. (2004). Particle flux and food supply to a seamount cold-water coral community Galicia Bank, NW Spain. *Marine Ecology Progress Series* 277: 13–23.

Duineveld, G.C.A.; Lavaleye, M.S.S.; Bergman, M.J.N.; de Stigter, H.; Mienis, F. (2007). Trophic structure of a coldwater coral mound community (Rockall Bank, NE Atlantic) in relation to the near-bottom particle supply and current regime. *Bulletin Marine Science* 81:449–467.

Dunstan, P.K.; Foster, S.D.; Darnell, R. (2011a). Model based grouping of species across environmental gradients. *Ecological Modelling* 222: 955–963.

Dunstan, P.K.; Althaus, F.; Williams, A.; Bax, N.J. (2011b). Characterising and predicting benthic biodiversity for conservation planning in deepwater environments. *PLoS ONE* 7(5): e36558. doi:10.1371/journal.pone.0036558

Elith, J.; Graham, C.H. (2009). Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* 32: 66–77.

Elith, J.; Leathwick, J.R. (2011). Boosted Regression Trees for ecological modelling. http://cran.r-project.org/web/packages/dismo/vignettes/brt.pdf. 22 p.

Elith, J.; Graham, C.H.; Anderson, R.P.; Dudik, M.; Ferrier, S.; Guisan, A.; Hijmans, R.J.; Huettmann, F.; Leathwick, J.R.; Lehmann, A.; Li, J.; Lohmann, L.G.; Loiselle, B.A.; Manion, G.; Moritz, C.; Nakamura, M.; Nakazawa, Y.; Overton, J.M.; Peterson, A.T.; Phillips, S.J.; Richardson, K.; Scachetti-Pereira, R.; Schapire, R.E.; Soberon, J.; Williams, S.; Wisz, M.S.; Zimmermann, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29: 129–151.

Elith, J.; Leathwick, J.R.; Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77:802–813.

Elith, J.; Phillips, S.J.; Hastie, T.; Dudík, M.; Chee, Y.E.; Yates, C.J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43–57.

FAO (2009). International Guidelines for the Management of Deep-Sea Fisheries in the High-Seas. Food and Agriculture Organisation of the United Nations, Rome, Italy, 73 p.

Garcia, H.E.; Locarnini, R.A.; Boyer, T.P.; Antonov, J.I. (2006). World Ocean Atlas 2005, Volume 4: Nutrients (phosphate, nitrate, silicate). S Levitus, ed. NOAA Atlas NESDIS 64, U.S. Government Printing Office, Washington DC. 396 p.

Genin, A.; Noble, M.; Lonsdale, P.F. (1989). Tidal currents and anticyclonic motions on two North Pacific seamounts. *Deep-Sea Research* 36:1803–1815.

Gorman, R.M.; Laing, A.K. (2000). A long-term wave hindcast for the New Zealand coast. 6th International workshop on wave hindcasting and forecasting. Monterey, California.

Guinotte, J.M.; Bartley, J.D.; Iqbal, A.; Fautin, D.G; Buddemeier, R.W. (2006). Modeling habitat distribution from organism occurrences and environmental data: case study using anemonefishes and their sea anemone hosts. *Marine Ecology Progress Series* 316: 269–283.

Guinotte, J.M.; Davies, A.J. (2012). Predicted deep-sea coral habitat suitability for the U.S. West Coast. Report to NOAA Deepsea Coral Program.

Guinan, J.; Grehan, A.J.; Dolan M.F.J.; Brown, C. (2009). Quantifying relationships between video observations of cold-water coral cover and seafloor features in Rockall Trough, west of Ireland. *Marine Ecology Progress Series* 375:125–138.

Guisan, A.; Lehmann, A.; Ferrier, S.; Austin, M.; Overton, J.M.C.C.; Aspinall, R.; Hastie, T. (2006). Making better biogeographical predictions of species distributions. *Journal of Applied Ecology* 43:386–392.

Hadfield, M.; Uddstrom, M.; Goring, D.; Gorman, R. et al. (2002). Physical variables for the New Zealand Marine Environment Classification System: development and description of data layers. NIWA Client Report CHC2002-043. NIWA, Wellington.

Hamann, A.; Aitken, S.N. (2013). Conservation planning under climate change: accounting for adaptive potential and migration capacity in species distribution models. *Diversity and Distributions* 19: 268–280.

Hanley, J.A.; McNeil, B.J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36.

Howell, K.L.; Holt, R.; Endrino, I.P.; Stewart, H. (2011). When the species is also a habitat: comparing the predictively modelled distributions of *Lophelia pertusa* and the reef habitat it forms. *Biological Conservation* 144: 2656–2665.

Leathwick, J.R.; Elith, J.; Francis, M.P.; Hastie, T.; Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series* 321:267–281.

Leathwick, J.R.; Elith, J.; Chadderton, W.L.; Rowe, R.; Hastie, T. (2008). Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *Journal of Biogeography* 35:1481–1497.

Leathwick, J.R.; Rowden, A.A.; Nodder, S.; Gorman, R.; Bardsley, S.; Pinkerton, M.; Baird, S.J.; Hadfield, M.; Currie, K.; Goh, A. (2009). Development of a benthic-optimised marine environment classification for waters within the New Zealand EEZ. Final Research Report prepared as part completion of Objective 5 of BEN200601 for the Ministry of Fisheries. (Unpublished report held by Ministry for Primary Industries.) 52 p.

Levin, L.A.; Etter, R.J.; Rex, M.A.; Gooday, A.J.; Smith, C.R.; Pineda, J.; Stuart, C.T.; Hessler, R.R.; Pawson, D. (2001). Environmental influences on regional deep-sea species diversity. *Annual Review of Ecology and* Systematics 32:51–93.

Li, J.; Heap, A.D.; Potter, A.; Haung, Z.; Daniell, J.J. (2011). Can we improve the spatial predictions of seabed sediments? A case study of spatial interpolation of mud content across the south west Australian margin. *Continental Shelf Research* 31: 1365–1376.

Lobo, J.M.; Jimenez-Valverde, A.; Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145–151.

Long, B.G.; Pitcher, C.R.; Bode, L.; Mason, I. (1997). Seabed Current Stress as a Predictor of the Distribution and Abundance of Epibenthos in Torres Strait. CSIRO Division of Marine Research Report MR-GIS 97/6.

Lutz, M.J; Caldeira, K.; Dunbar, R.B.; Behrenfeld, M.J. (2007). Seasonal rhythms of net primary production and particulate organic carbon flux to depth describe the efficiency of biological pump in the global ocean. *Journal of Geophysical Research* 112: C10011.

Ministry of Fisheries (2008). Bottom Fishery Impact Assessment: bottom fishing activities by New Zealand vessels fishing in the High Seas in the SPRFMO area during 2008 and 2009. (Unpublished report available from Ministry for Primary Industries, Wellington.) 109 p.

Morato, T.; Pitcher, T.J.; Clark, M.R.; Menezes, G.; Tempera, F.; Porteiro, F.; Giacomello, E.; Santos, R.S. (2010). Can we protect seamounts for research? A call for conservation. *Oceanograph*y 23(1): 190–199.

Orr, J.C.; Fabry, V.J.; Aumont, O.; Bopp, L.; Doney, S.C. et al. (2005) Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature* 437: 681–686.

Papeş, M.; Gaubert, P. (2007). Modelling ecological niches from low numbers of occurrences: assessment of the conservation status of poorly known viverrids (Mammalia, Carnivora) across two continents. *Diversity and Distributions* 13: 890–902.

Parker, S.J.; Penney, A.J.; Clark, M.R. (2009). Detection criteria for managing trawl impacts on vulnerable marine ecosystems in high seas fisheries of the South Pacific Ocean. *Marine Ecology Progress Series,* 397: 309–317.

Pearson, R.G; Raxworthy, C.J.; Nakamura, M.; Peterson, A.T. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102–117.

Penney, A.J.; Parker, S.J.; Brown, J.H. (2009). Protection measures implemented by New Zealand for vulnerable marine ecosystems in the South Pacific Ocean. *Marine Ecology Progress Series,* 397: 341–354.

Peterson, A.T.; Papes, M.; Soberon, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* 213: 6372.

Pinkerton, M.H.; Moore, G.F.; Lavender, S.J.; Gall, M.P. et al. (2006). A method for estimating inherent optical properties of New Zealand continental shelf waters from satellite ocean colour measurements. *New Zealand Journal of Marine and Freshwater Research* 40: 227–247.

Pitcher, C.R.; Lawton, P.; Ellis, N.; Smith, S.J.; Incze, L.S.; Wei, C.-L.; Greenlaw, M.E.; Wolff, N.H.; Sameoto, J.A.; Snelgrove, P.V.R. (2012). Exploring the role of environmental variables in shaping patterns of seabed biodiversity composition in regional-scale ecosystems. *Journal of Applied Ecology* 49: 670–679.

Phillips, S.J.; Anderson, R.P.; Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190: 231–259.

Phillips, S.J.; Dudik, M.; Elith, J.; Graham, C.H.; Lehmann, A.; Leathwick, J.; Ferrier, S. (2009). Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, 19:181–197.

Phillips, S.J.; Dudik, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161–175.

R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna. http://www.R-project.org.

Rengstorf, A.M.; Grehan, A.;Yesson, C.; Brown, C. (2012) Towards high resolution habitat suitability modelling of vulnerable marine ecosystems in the deep-sea: resolving terrain attribute dependencies. *Marine Geodesy* 35: 343–361.

Ridgeway, G. (2006). 'gbm' generalized boosted regression models, R Package Version 1.5. Accessed July 2008.

Riordan, E.C.; Rundel, P.W. (2009). Modelling the distribution of a threatened habitat: the California sage scrub. *Journal of Biogeography* 36: 2176–2188.

Rowden, A.A.; Oliver, M.; Clark, M.R.; MacKay, K. (2008). New Zealand's "SEAMOUNT" database: recent updates and its potential use for ecological risk assessment. Aquatic Environment and Biodiversity Report No. 27. 49 p.

Snelgrove, P.V.R.; Butman, C.A. (1994). Animal-sediment relationships revisited: cause versus effects. *Oceanography and Marine Biology: an Annual Review* 32: 111–177.

SPRFMO (2007). Report of the Science Working Group. 4th SPRFMO Meeting, Noumea, New Caledonia, 4–7 September 2007, 12 p plus appendices.

Steinacher, M.; Joos, F.; Frölicher, T.L.; Plattner, G-K.; Doney, S.C. (2009). Imminent ocean acidification projected with the NCAR global coupled carbon cycleclimate model. *Biogeosciences* 6: 515–533.

Swets, J.A. (1988). Measuring the accuracy of diagnostic systems. Science 240: 1285-1293. doi:10.1126/science.3287615. PubMed: 3287615.

Tittensor, D.P.; Baco, A.R.; Brewin, P.E.; Clark, M.R.; Consalvey, M.; Hall-Spencer, J.; Rowden, A.A.; Schlacher, T.; Stocks, K.I.; Rogers, A.D. (2009). Predicting global habitat suitability for stony corals on seamounts. *Journal of Biogeography*, 36: 1111–1128.

Tittensor, D.P.; Baco, A.R.; Hall-Spencer, J.M.; Orr, J.C.; Rogers, A.D. (2010). Seamounts as refugia from ocean acidification for cold-water stony corals. *Marine Ecology*, 31(Suppl. 1):212–225.

Thiem, Ø.; Ravagnan, E.; Fosså, J.H.; Bersten, J. (2006). Food supply mechanisms for coldwater corals along a continental shelf edge. *Journal of Marine Systems* 60: 207–219.

Thistle, D. (2003). The deep-sea floor: an overview. In: Ecosystems of the Deep Ocean, Ed. Tyler, P.A. Elsevier, Amsterdam. pp 5–37.

Torres, L.G.; Smith, T.D.; Sutton, P.; MacDiarmid, A.; Bannister, J.; Miyashita, T. (2013). From exploitation to conservation: habitat models using whaling data predict distribution patterns and threat exposure of an endangered whale. *Diversity and Distributions*. DOI: 10.1111/ddi.12069

Tracey, D.; Clark, M.; Consalvey, M.; Hourigan, T. (2010). Report on the Joint Science and Technology Commission Meeting (JCM) Ocean and Marine Sciences Workshop. Wellington, New Zealand, 25 – 26 January 2010. National Marine Fisheries Service (NOAA) & National Institute of Water & Atmospheric Research (NIWA), 59 p.

Tracey, D.M.; Rowden, A.A.; Mackay, K.A.; Compton, T. (2011a). Habitat-forming cold-water corals show affinity for seamounts in the New Zealand region. *Marine Ecology Progress Series* 430: 1–22.

Tracey, D.; Baird, S.J.; Sanders, B.; Smith, M.H. (2011b). Distribution of protected corals in relation to fishing effort and assessment of accuracy of observer identification. NIWA Client Report WLG2011-33 prepared for Marine Conservation Services (MCS) Department of Conservation/Te Papa Atawhai, 70p. http://www.doc.govt.nz/documents/conservation/marine-and-coastal/marine-conservation-services/mcsint2010-03-coral-bycatch-final-report.PDF

Tyberghein, L.; Verbruggen, H.; Pauly, K.; Troupin, C.; Mineur, F.; De Clerck, O. (2011). Bio-ORACLE: a global environmental dataset for marine species distribution modelling. *Global Ecology and Biogeography* DOI: 10.1111/j.1466-8238.2011.00656.x

Walters, R.A.; Goring, D.G.; Bell, R.G. (2001) Ocean tides around New Zealand. *New Zealand Journal of Marine and Freshwater Research* 35:567–579.

Warwick, R.M.; Uncles, R.J. (1980). Distribution of Benthic Macrofauna Associations in the Bristol Channel in Relation to Tidal Stress. *Marine Ecology Progress Series* 3: 97–103.

Williams, A.; Schlacher, T.A.; Rowden, A.A.; Althaus, F.; Clark, M.R.; Bowden, D.A.; Stewart, R.; Bax, N.J.; Consalvey, M.; Kloser, R.J. (2010). Seamount megabenthic assemblages fail to recover from trawling impacts. *Marine Ecol*ogy 31 (Suppl 1): 183–199.

Wood, A.C.L.; Rowden, A.A.; Compton, T.J.; Gordon, D.P.; Probert, P.K. (2013) Habitat-forming bryozoans in New Zealand: their known and predicted distribution in relation to broad-scale environmental variables and fishing effort. PLoS ONE 8(9): e75160. doi:10.1371/journal.pone.0075160

Woodby, D.; Carlile, D.; Hulbert, L. (2009). Predictive modeling of coral distribution in the Central Aleutian Islands, USA. *Marine Ecology Progress Series* 397:227–240.

Yesson, C.; Taylor, M.L.; Tittensor, D.P.; Davies, A.J.; Guinotte, J.; Baco, A.; Black, J.; Hall-Spencer, J.M.; Rogers, A.D. (2012). Global habitat suitability of cold-water octocorals. *Journal of Biogeography* doi:10.1111/j.1365-2699.2011.02681.x

TABLES

Table 1: List of 10 taxonomic groups assessed as being vulnerable to bottom trawl fishing in the South Pacific Ocean (from Parker et al. 2009).

Taxonomic level	Taxonomic name	Common name
Vulnerable taxa		
Phylum	Porifera	Sponges
Order	Actiniaria	Anemones
Order	Alcyonacea	Soft corals
Order	Gorgonacea	Sea fans
Order	Pennatulacea	Sea pens
Order	Scleractinia	Stony corals
Order	Antipatharia	Black corals
Family	Stylasteridae	Hydrocorals
Habitat indicators		
Class	Crinoidea	Sea lilies
Order	Brisingida	Armless stars

Table 2: Number of data records for each VME indicator taxon (listed alphabetically) that were available for use in models (those with complete latitude, longitude, depth data).

Taxon	*n* records for use in models
Actinaria	8032
Alcyonacea	712
Antipatharia	1025
Brisingida	505
Crinoidea	558
Gorgonacea	2401
Pennatulacea	936
Porifera	9187
Scleractinia	4738
Stylasteridae	1061

Table 3: Description of the ‘regional’ environmental variables used in the model analysis, where the variable name, description, units and original source of the data are shown.

Variable	Description	Unit	Source
Bottom current speed	Mean current speed at the seabed	$dm\ s^{-1}$	Gorman unpublished – derived from Gorman & Laing (2000)
Tidal current speed	Depth averaged maximum tidal current speed	$m\ s^{-1}$	Gorman unpublished – derived from Walters et al. (2001), Hadfield et al. (2002)
Dynamic topography	Mean sea surface height above geoid reference	m	Hadfield unpublished - derived from AVISO and corrected from geophysical effects taken for the NZ region
SST gradient	Annual mean spatial gradient of sea surface temperature (SST)	$^{o}C\ km^{-1}$	Hadfield unpublished - derived from NIWA SST archive; see Uddstrom & Oien (1999), Hadfield et al. (2002)
Seamount	Seamounts	-	New Zealand SEAMOUNT database (Rowden et al. 2008)
Temperature	Mean annual water temperature at the seafloor	°C	CARS 2009 (www.cmar.csiro.au/cars)
Depth	Depth at the seafloor as interpolated from contours generated from various sources (e.g. multi-beam and single-beam echo sounders)	m	Mackay unpublished – derived from NIWA bathymetric data archive
Slope	Seafloor slope (1 km scale)	degrees	Mackay unpublished – derived from analysis of NIWA regional bathymetric data
Primary productivity	Mean annual net primary productivity in surface water	$mgC\ m^{-2}d^{-1}$	Behrenfield & Falkowski (1997); www.science.oregonstate.edu/ocean.productivity/index.php
Dissolved organic matter	Ocean colour indicative of dissolved organic matter	aDOM (443) $m^{-}1$	Pinkerton unpublished – derived using modified Case 2 corrections and algorithms applied to ocean colour remote-sensed data (Pinkerton et al. 2006)
POC flux	Particulate organic carbon (POC) flux to the seafloor	$g\ Corg\ m^{-2}\ yr^{-1}$	Lutz et al. (2007)

Table 4: Description of the 'global' environmental variables used in the model analysis, where the variable name, description, units and original source of the data are shown (see Davies & Guinotte (2011) for further detail).

Variable	Description	Unit	Original data source
Temperature	Annual mean temperature at the seafloor	°C	Boyer et al. (2005)
Depth	Water depth from SRTM30 bathymetry	m	Becker et al. (2009)
Dissolved oxygen	Annual mean dissolved oxygen at the seafloor	$ml\ l^{-1}$	Garcia et al. (2006)
Slope	Mean slope of SRTM30 bathymetry	$m\ m^{-1}$	derived from Becker et al. (2009)
Salinity	Annual mean salinity at the seafloor	PSU	Boyer et al. (2005)
POC flux	Particulate organic carbon flux to the seafloor	$g\ C_{org}\ m^{-2}\ yr^{-1}$	Lutz et al. (2007)
Calcite	Annual mean calcite concentration at the seafloor	Ω_{calc}	Steinacher et al. (2009)
Aragonite	Annual mean aragonite concentration at the seafloor	Ω_{arag}	Orr et al. (2005)
Silica	Annual mean silicate at the seafloor	$\mu mol\ l^{-1}$	Garcia et al. (2006)

Table 5: Table showing the performance estimates for the Maxent models of VME indicator taxa using 'regional' environmental data (performance results based on ten-fold cross validation, 500 iterations). (n= number of records for each model; see main text for explanation of Omission Rate, Test Gain and AUC).

.

Taxon	n	Omission Rate (%)	Test Gain	AUC
Actinaria	7318	1.1	2.243	0.974
Alcyonacea	600	1.8	2.644	0.971
Antipatharia	831	1.9	2.959	0.978
Brisingida	453	2.1	3.243	0.975
Crinoidea	452	3.2	2.776	0.970
Gorgonacea	2146	2.4	2.295	0.969
Pennatulacea	793	2.5	2.879	0.977
Porifera	8128	1.2	2.015	0.966
Scleractinia	4019	2.1	2.084	0.965
Stylasteridae	953	2.8	2.417	0.961

Table 6: Table showing the performance estimates of the BRT models for VME indicator taxa using 'regional' environmental data (n = number of records for VME indicator taxon used in model; see main text for explanation of % deviance explained and AUC).

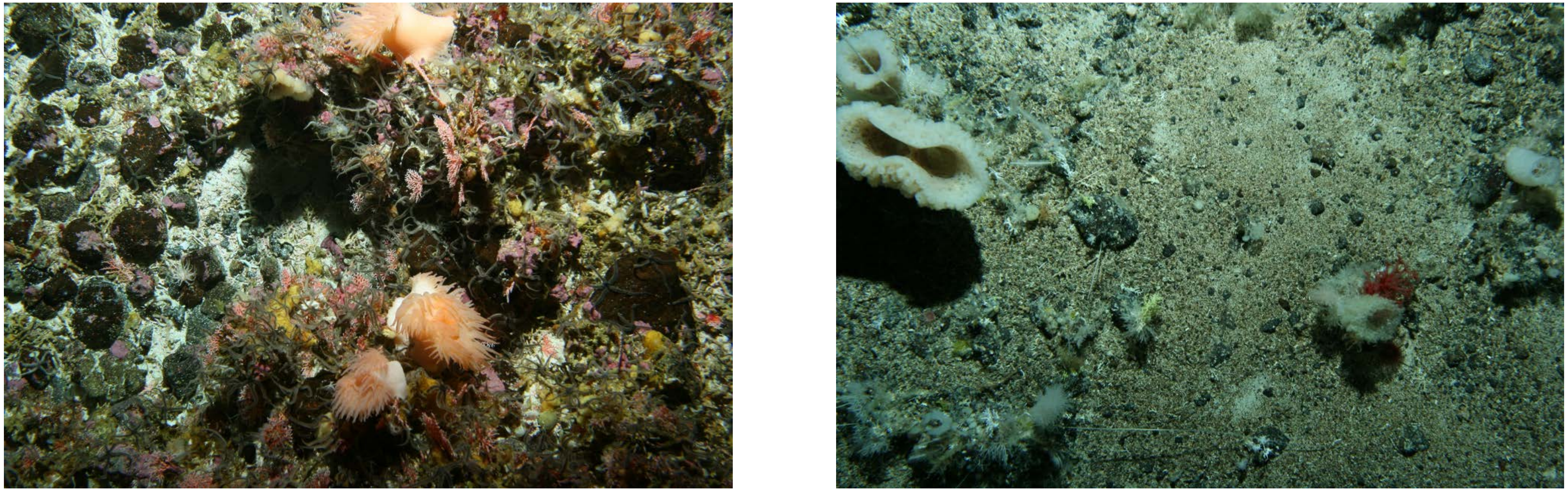
Taxon	n	% deviance explained	AUC
Actinaria	8032	0.42	0.88
Alcyonacea	712	0.23	0.77
Antipatharia	1025	0.51	0.91
Brisingida	505	0.36	0.82
Crinoidea	558	0.33	0.77
Gorgonacea	2401	0.39	0.87
Pennatulacea	936	0.30	0.83
Porifera	9187	0.33	0.85
Scleractinia	4738	0.34	0.85
Stylasteridae	1061	0.35	0.83

Table 7: Table showing the comparison of the area under the curve (AUC) measure of performance for Maxent models using 'regional' and 'global' environmental data. Maximum and minimum values are underlined, and the best performing model is marked with an * when the difference in the AUC value between models is more than 0.1.

Taxon	AUC (Global)	AUC (Regional)
Actinaria	<u>0.81</u>	0.98
Alcyonacea	0.97	0.98
Antipatharia	0.97	0.98
Brisingida	0.97	<u>0.99</u>
Crinoidea	0.96	0.97
Gorgonacea	0.96	0.97
Pennatulacea	0.95	0.98
Porifera	0.96	0.97
Scleractinia	0.97	0.97
Stylasteridae	0.97	0.97

FIGURES

Figure 1: Photographs of the seafloor showing examples of the VME indicator taxa of the South Pacific Ocean.

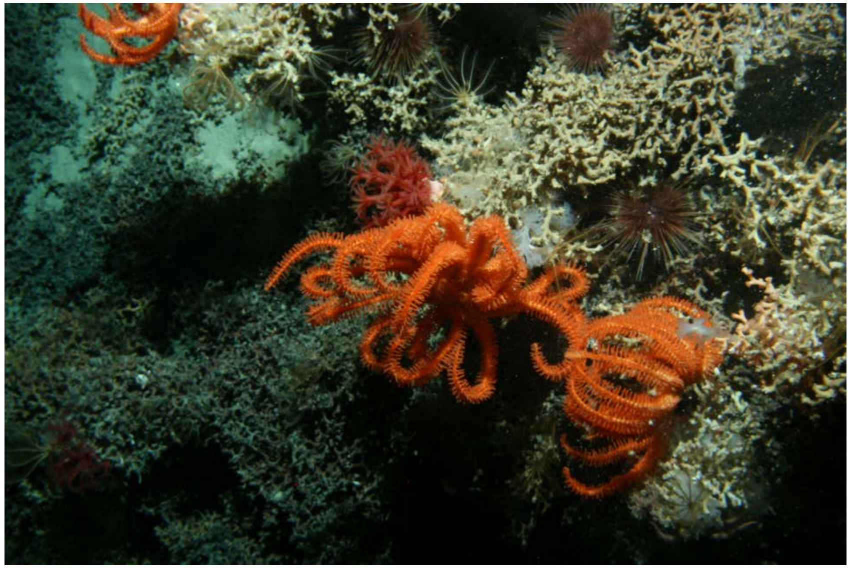


(a) Actinaria (orange-coloured organisms in centre of photo)

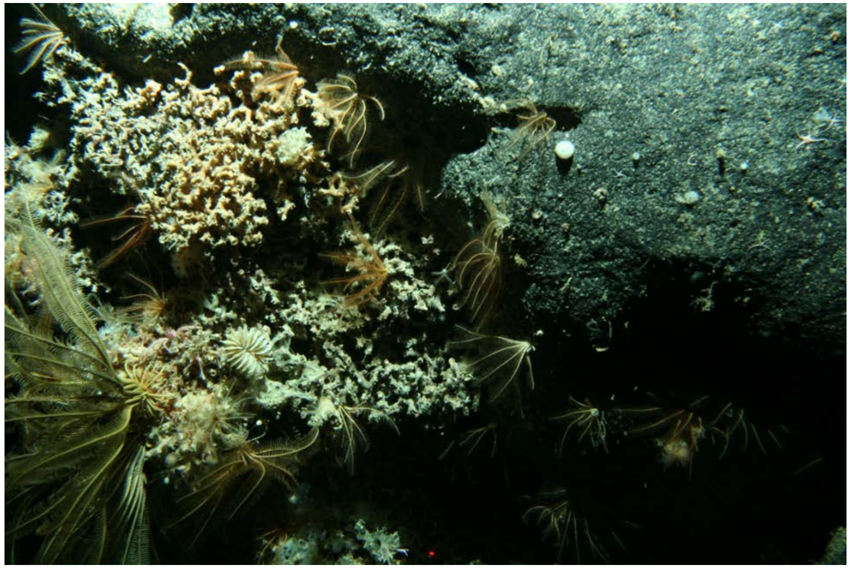
(b) Alcyonacea (brown-coloured organism attached to white organism)

(c) Antipatharia (brown-coloured organism)

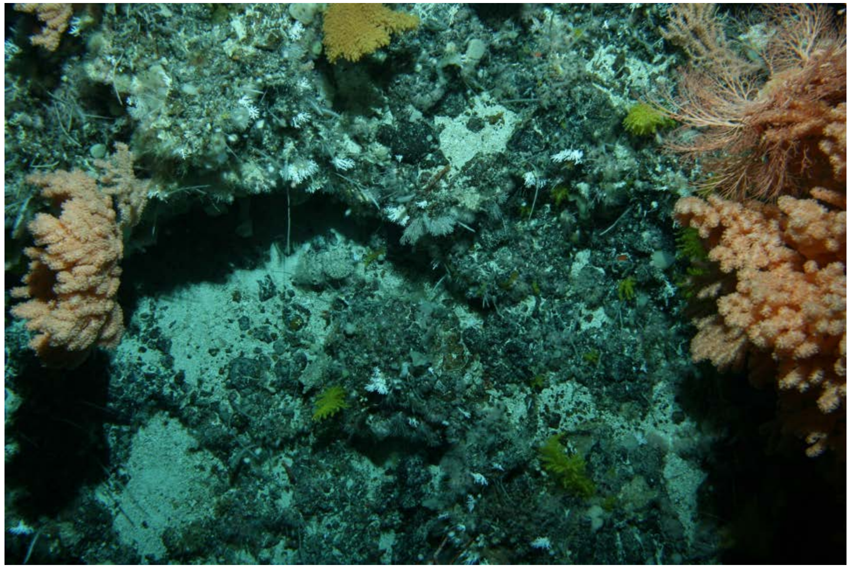
(d) Brisingida (orange-coloured organisms)



(e) Crinoidea (pale yellow-coloured organisms)

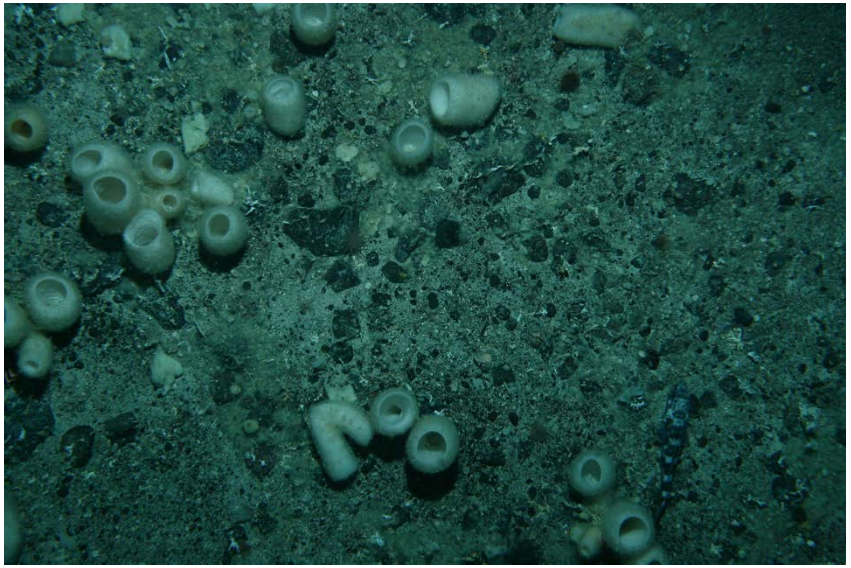


(f) Gorgonacea (pale orange-coloured organisms)

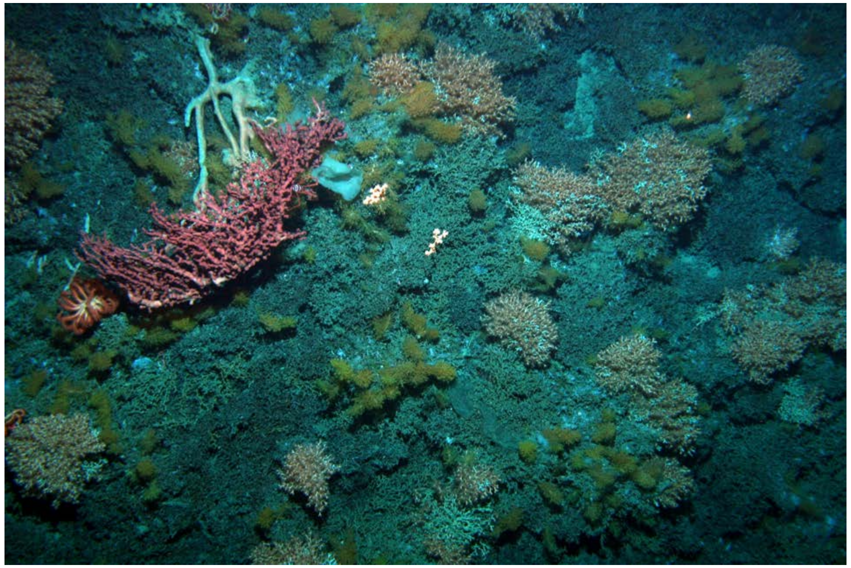


(g) Pennatulacea (pale purple-coloured organism)

(h) Porifera (white organisms)



(i) Scleractinia (white and pale pink-coloured branching organisms)



(j) Stylasteridae (white finely-branched organisms)

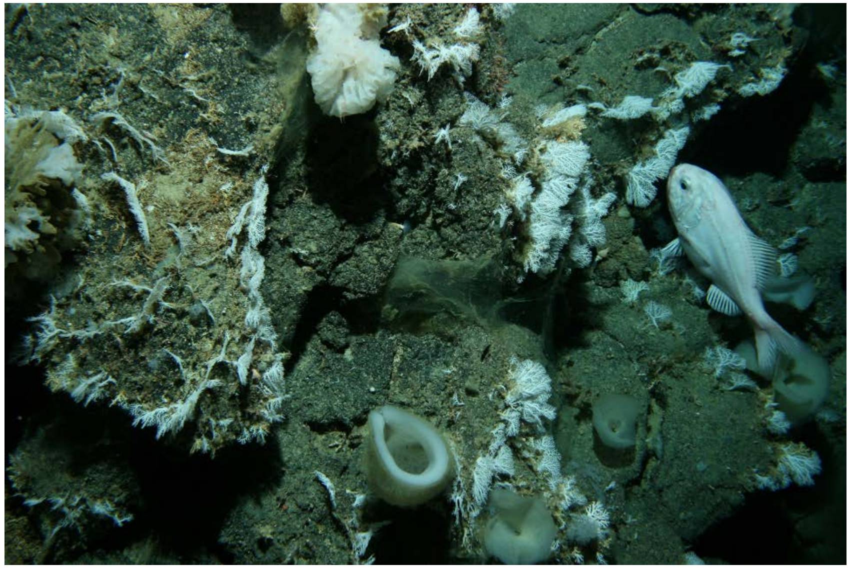


Figure 2: Map showing the SPRFMO area and the study area for the habitat suitability modelling.

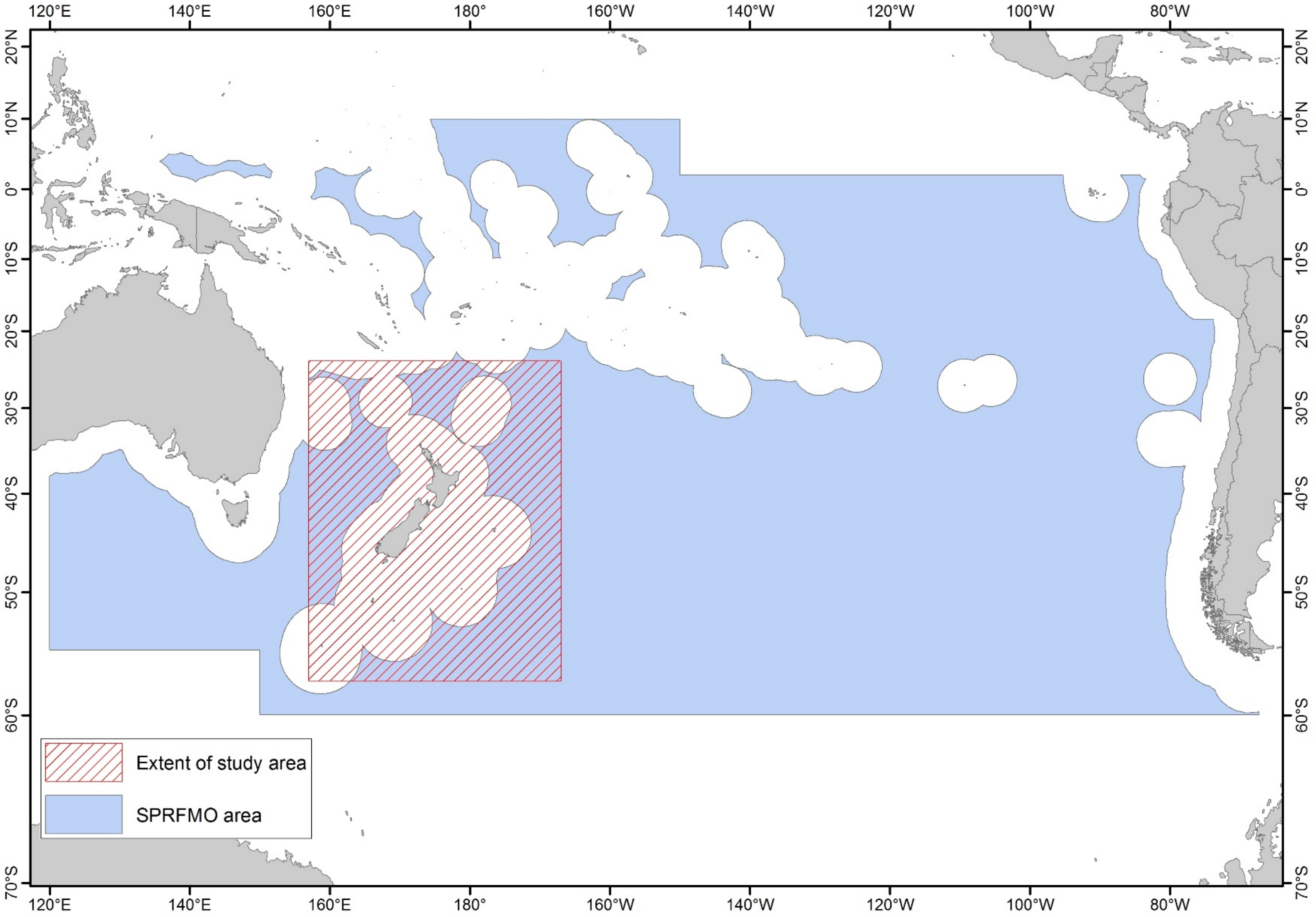
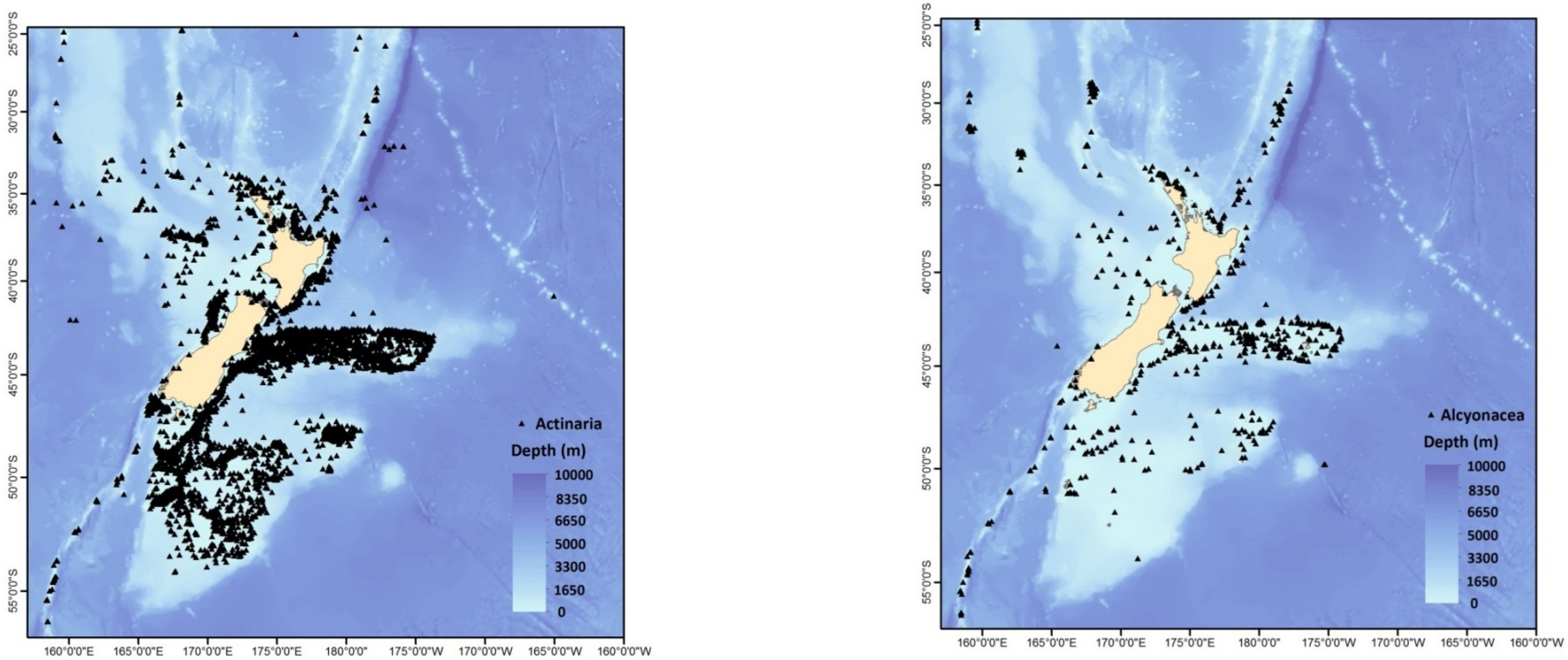


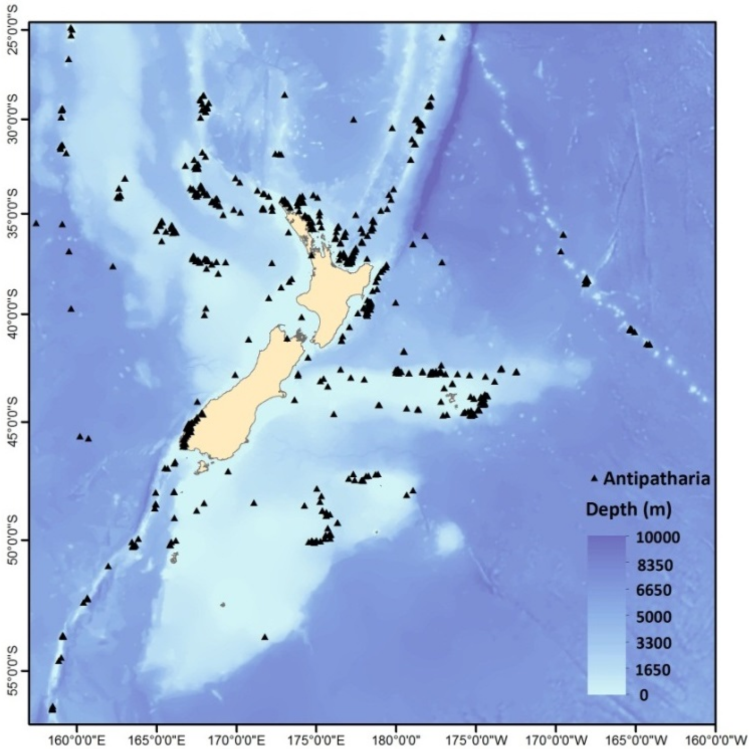
Figure 3: Maps showing the distribution of records for VME indicator taxa in the study area.



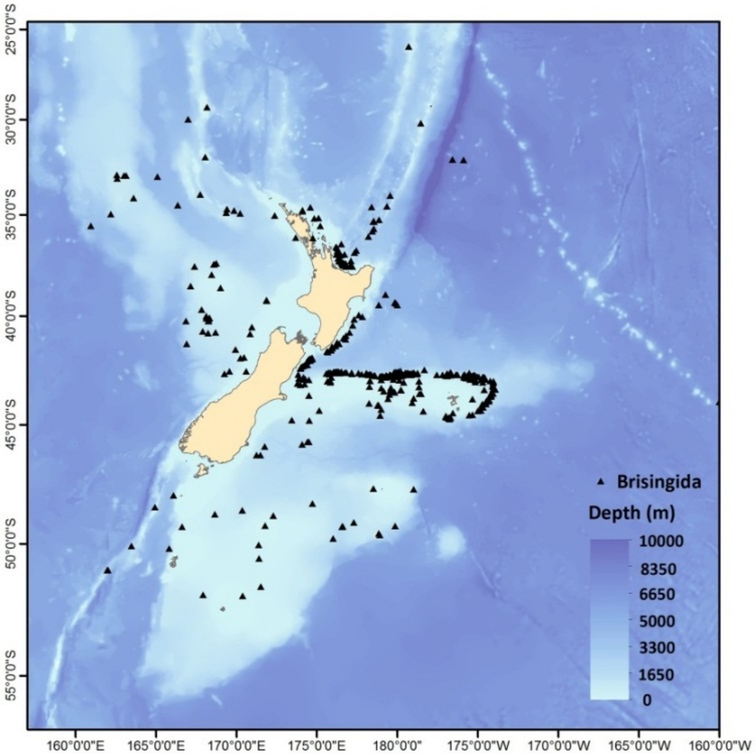
(a) Actinaria

(b) Alcyonacea

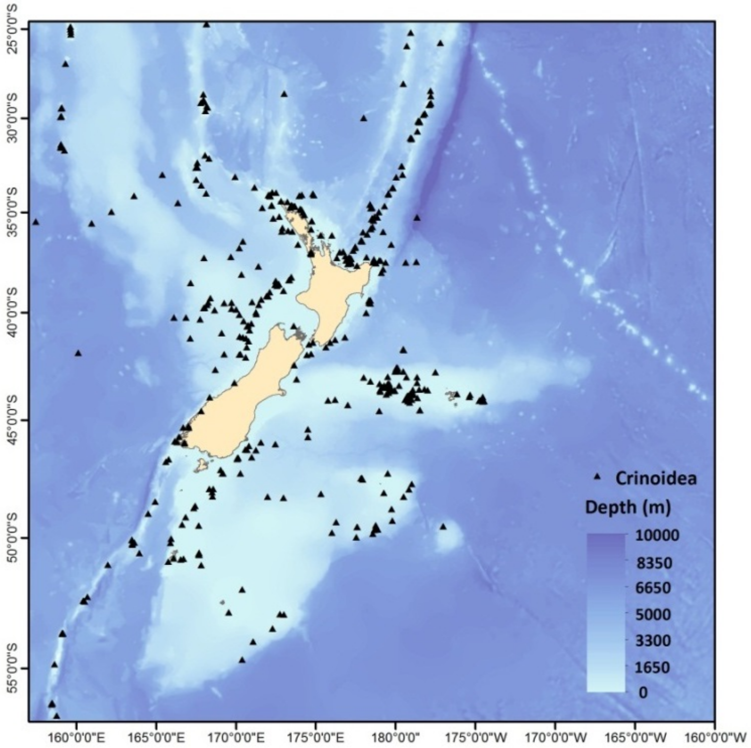
(c) Antipatharia



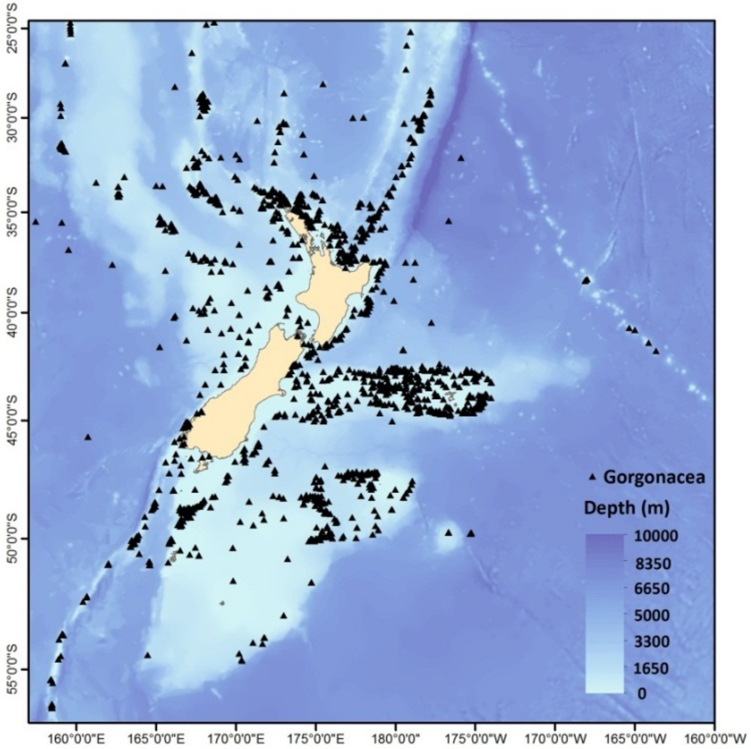
(d) Brisingida



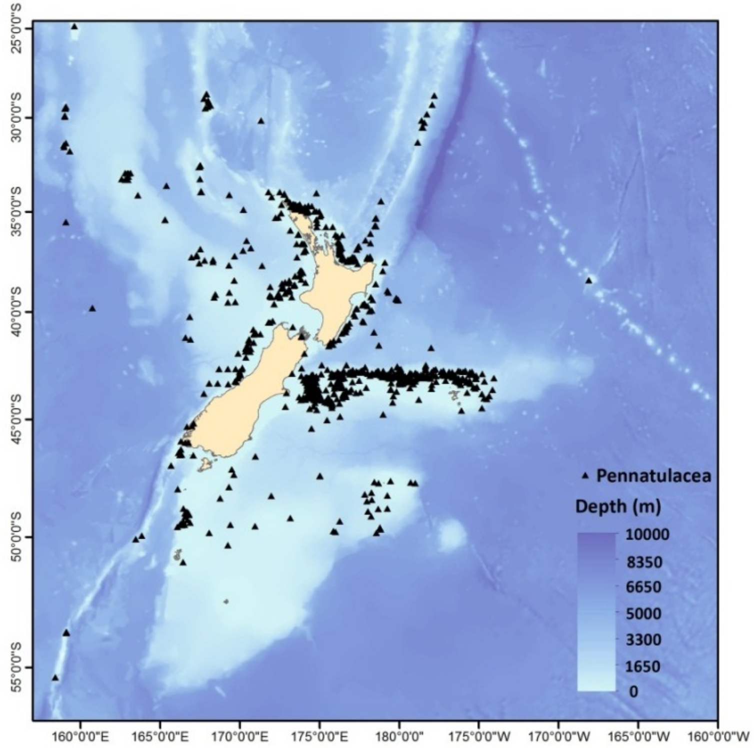
(e) Crinoidea



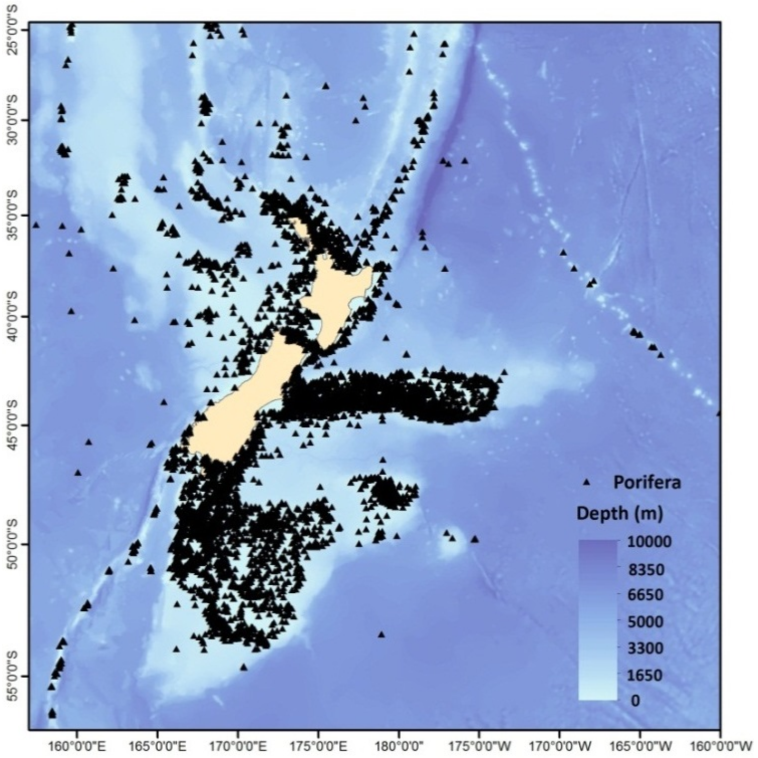
(f) Gorgonacea

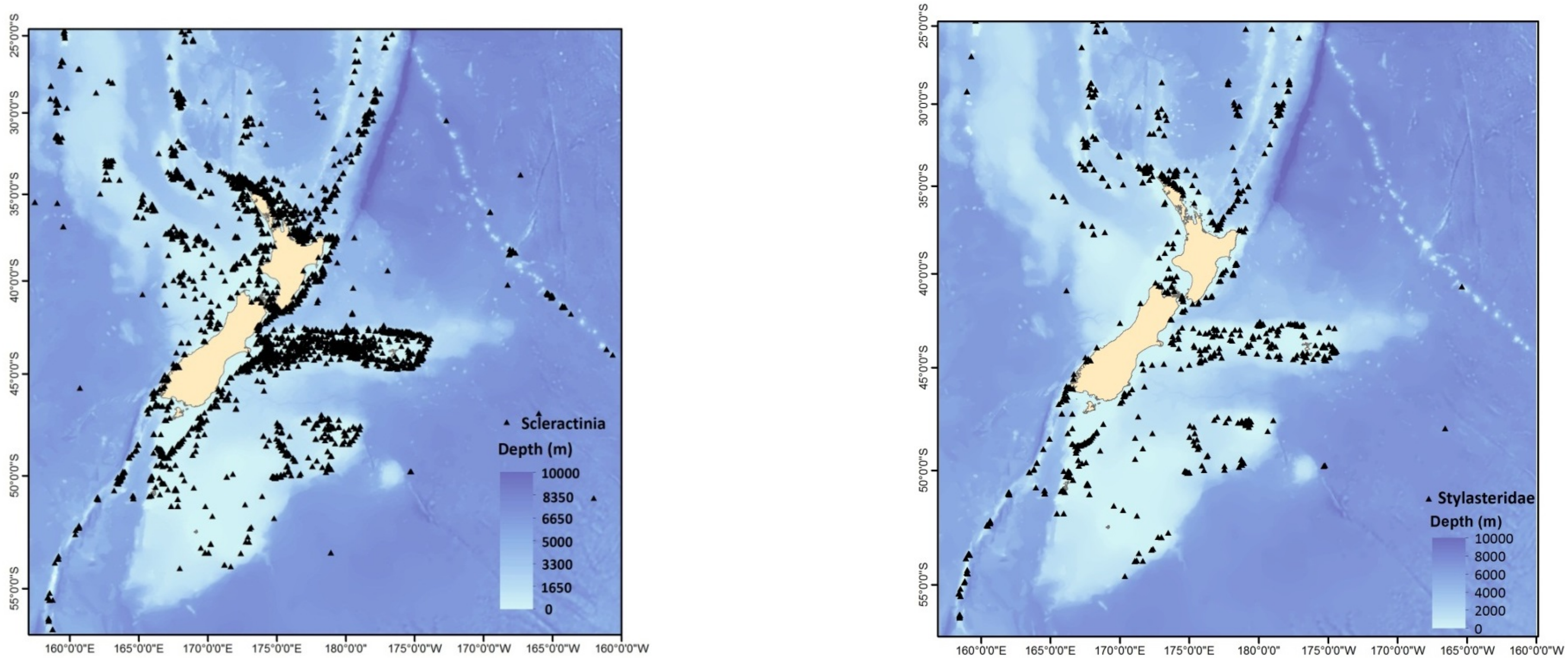


(g) Pennatulacea



(h) Porifera

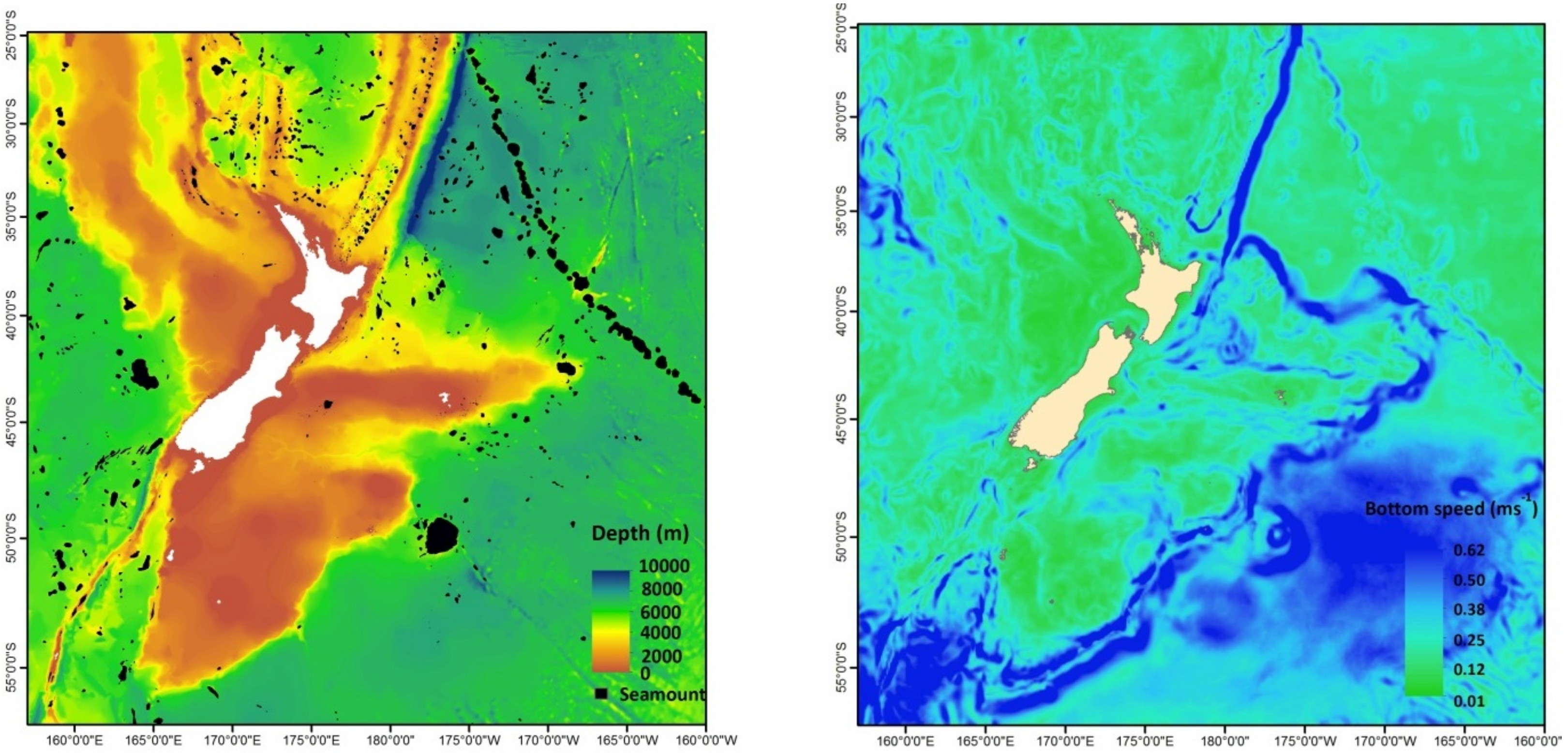


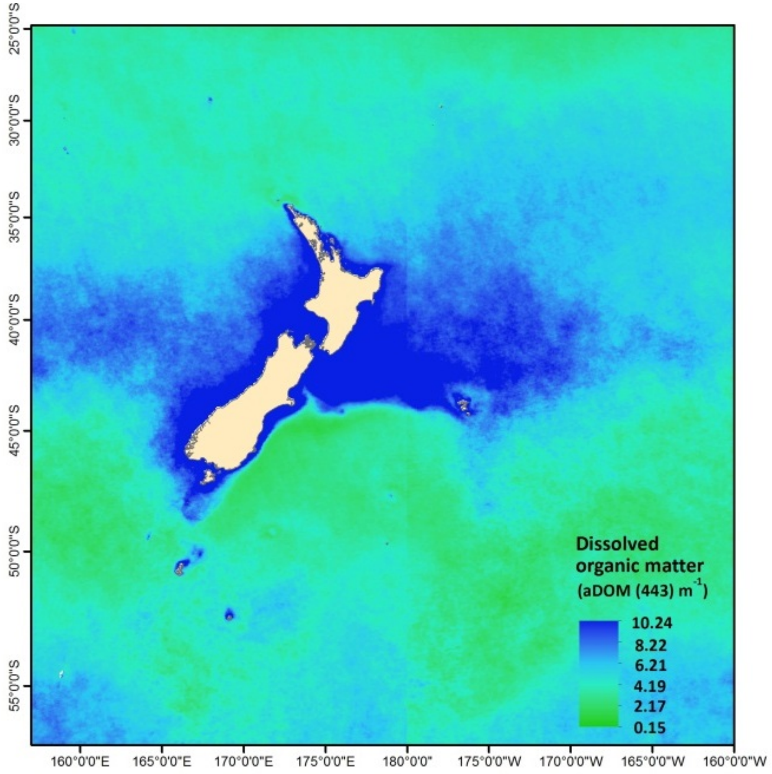


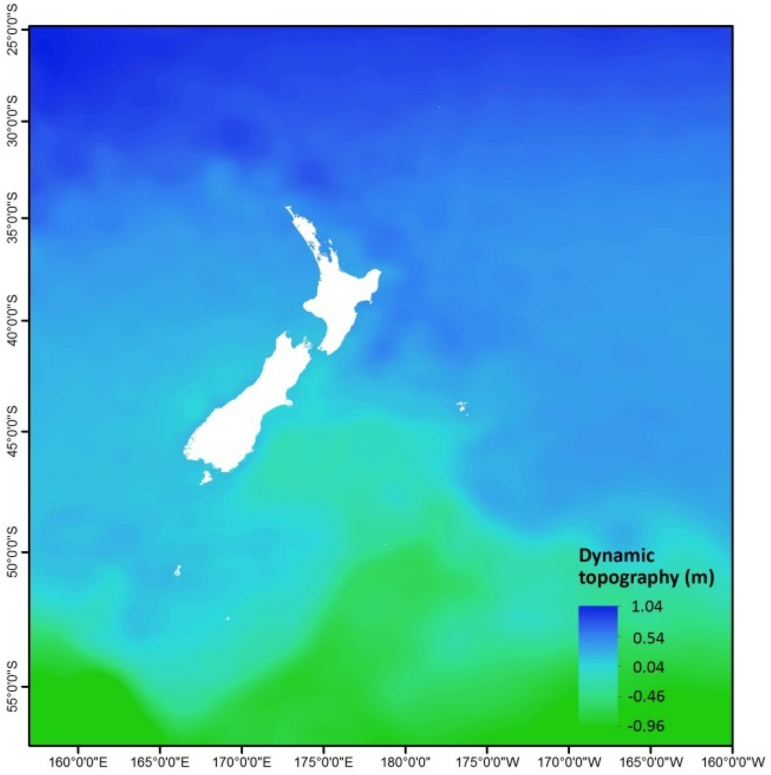
(i) Scleractinia

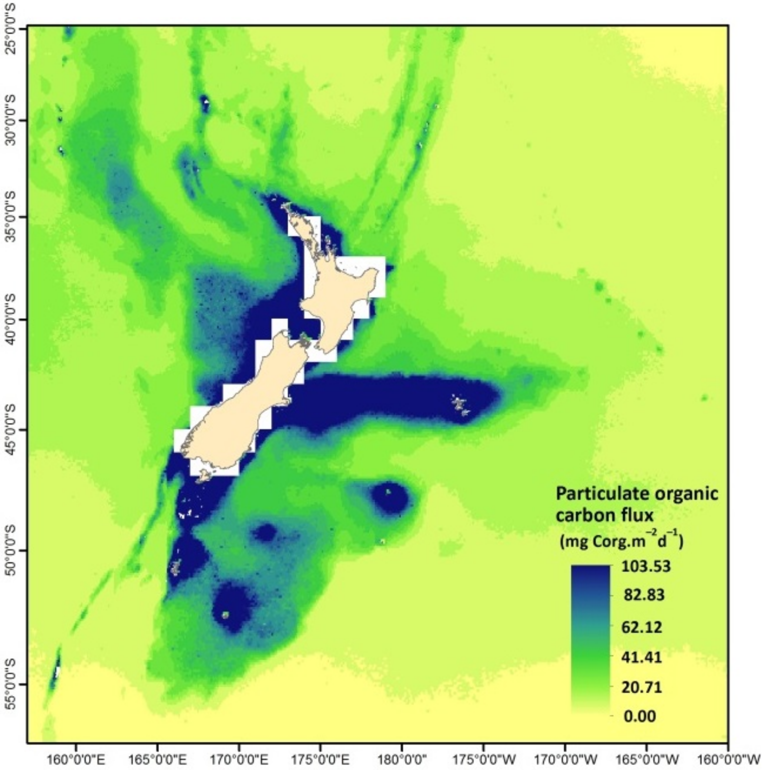
(j) Stylasteridae

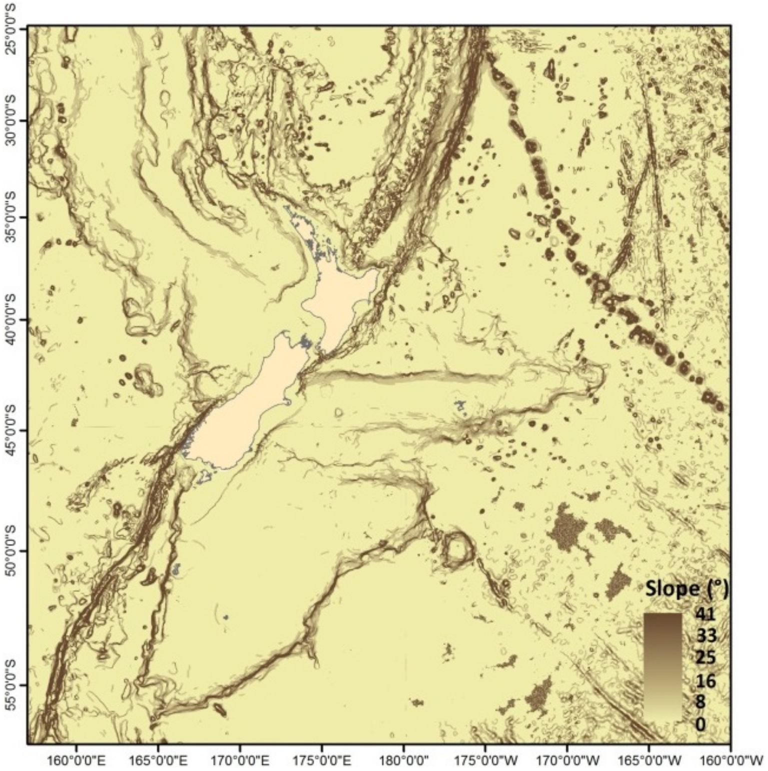
Figure 4: Maps showing the ‘regional’ environmental data layers used for the habitat suitability models (note: seamounts are shown on the depth data layer).

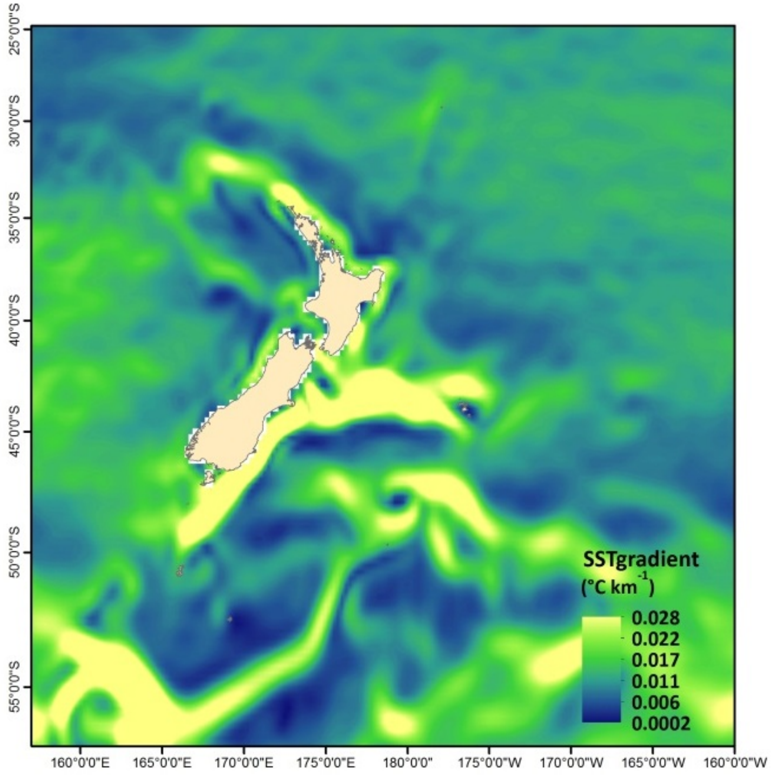


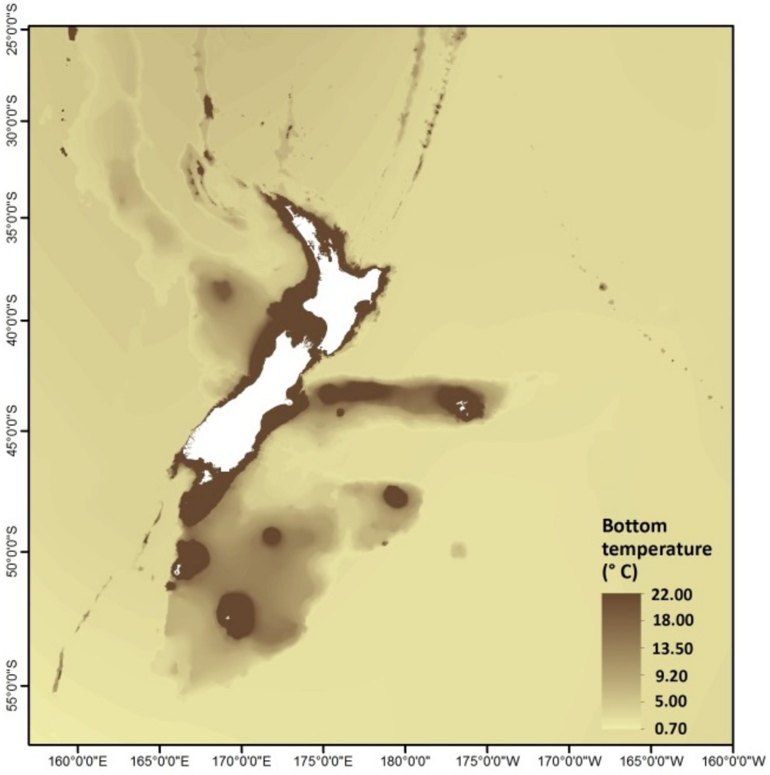

25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W
Dissolved organic matter
(aDOM (443) m^{-1})
10.24
8.22
6.21
4.19
2.17
0.15

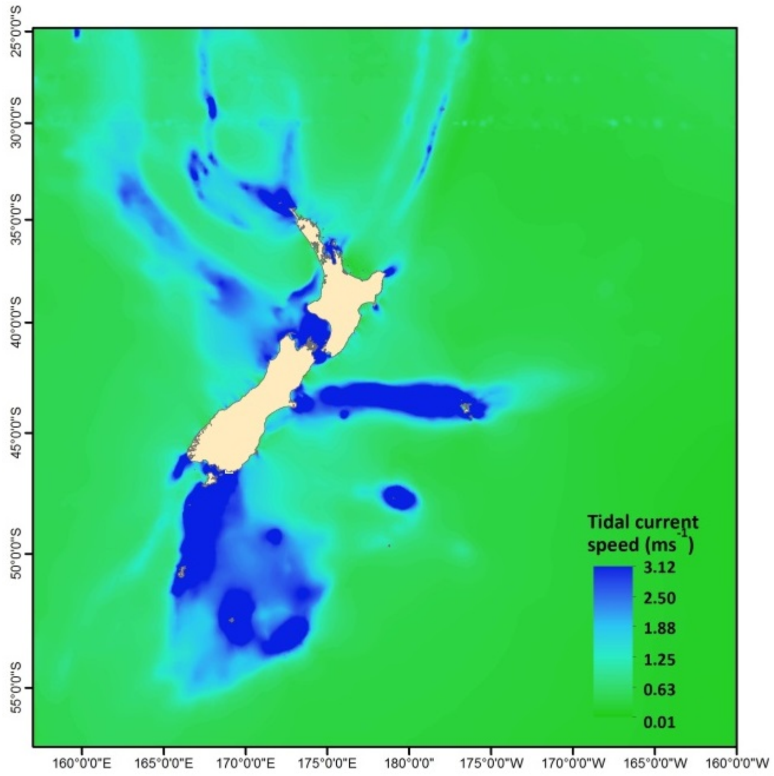

25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W
Dynamic topography (m)
1.04
0.54
0.04
-0.46
-0.96


Particulate organic carbon flux
(mg Corg.m^{-2}d^{-1})
103.53
82.83
62.12
41.41
20.71
0.00
25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W


Slope (°)
41
33
25
16
8
0
25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W


SSTgradient
(°C km^{-1})
0.028
0.022
0.017
0.011
0.006
0.0002
25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W


Bottom temperature
(° C)
22.00
18.00
13.50
9.20
5.00
0.70
25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W


Tidal current speed (ms⁻¹)
3.12
2.50
1.88
1.25
0.63
0.01
25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W

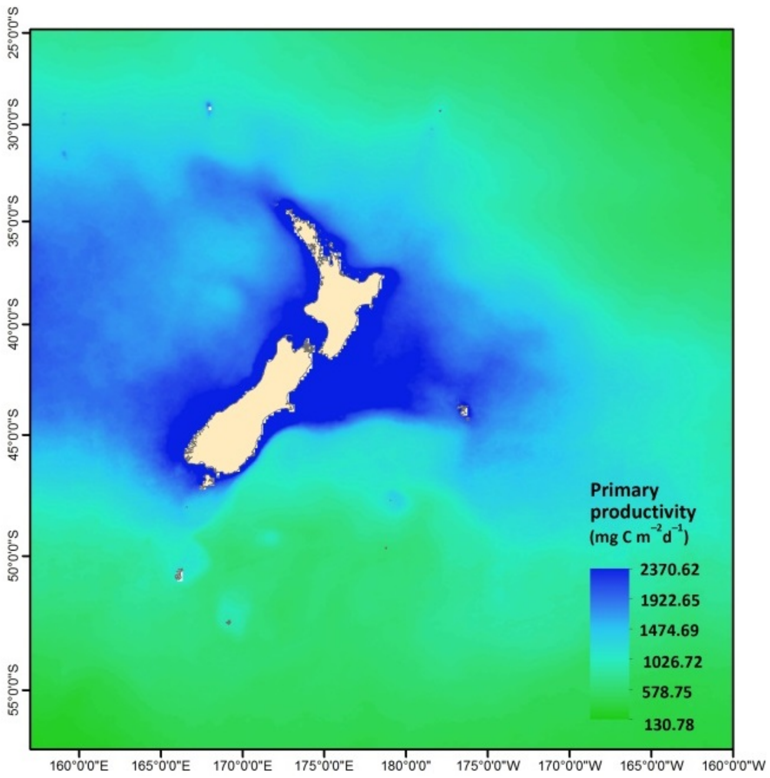
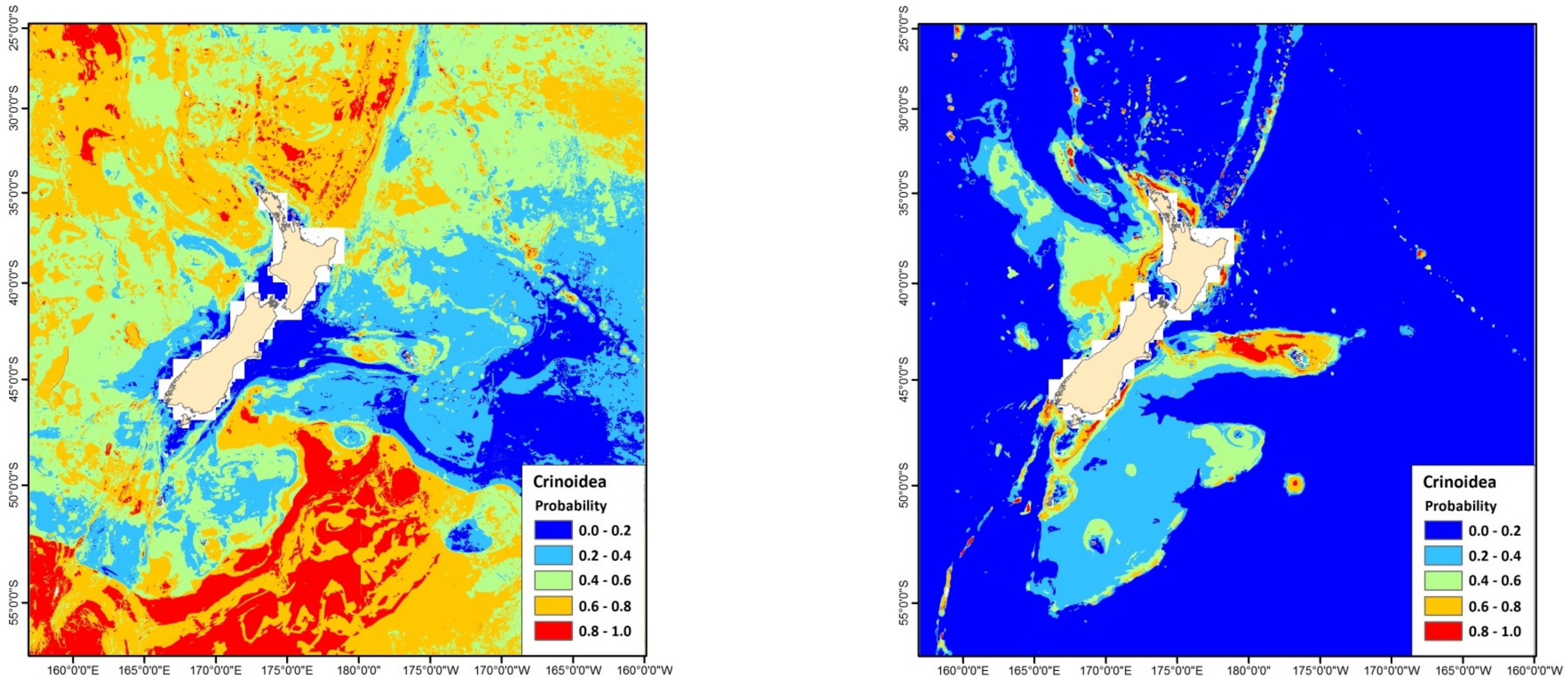

Primary productivity (mg C m⁻² d⁻¹)
2370.62
1922.65
1474.69
1026.72
578.75
130.78
25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W

Figure 5: BRT (left) and Maxent (right) habitat suitability outputs for example taxa showing (a) a contrasting prediction, and (b) a similar prediction. (key shows probability of habitat suitability from 0 – 1).

(a) Crinoidea – contrasting prediction



(b) Pennatulacea – similar prediction

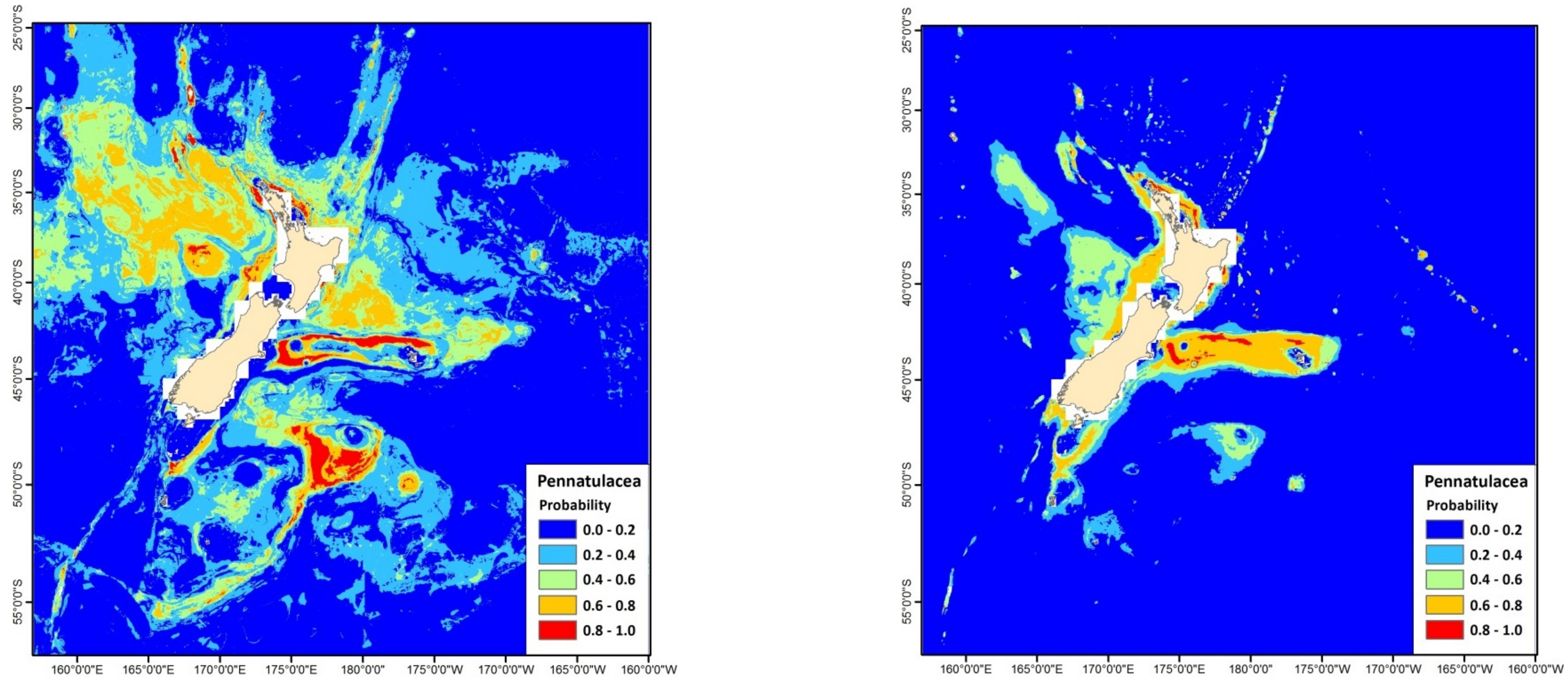


Figure 6: BRT (left) and Maxent (right) habitat suitability outputs for Scleractinia at the Graveyard Seamount Complex (northern Chatham Rise) showing the contrasting spatial variability of the predictions. Also shown is the number of presence records per 1 km^2 cell used for the BRT model (left), and the location of the individual presence records of which only one record per 1 km^2 cell was used for the Maxent model (right).

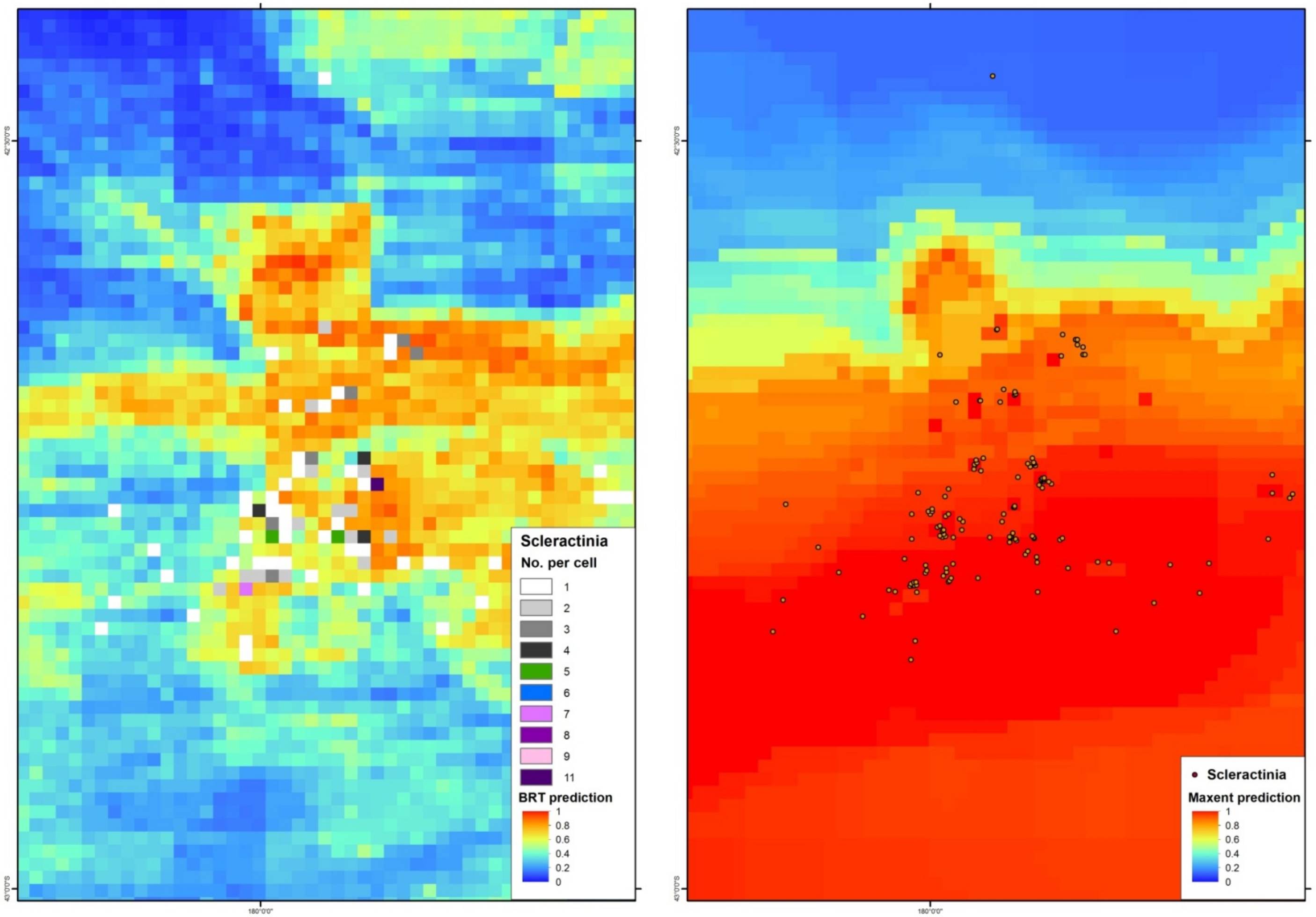
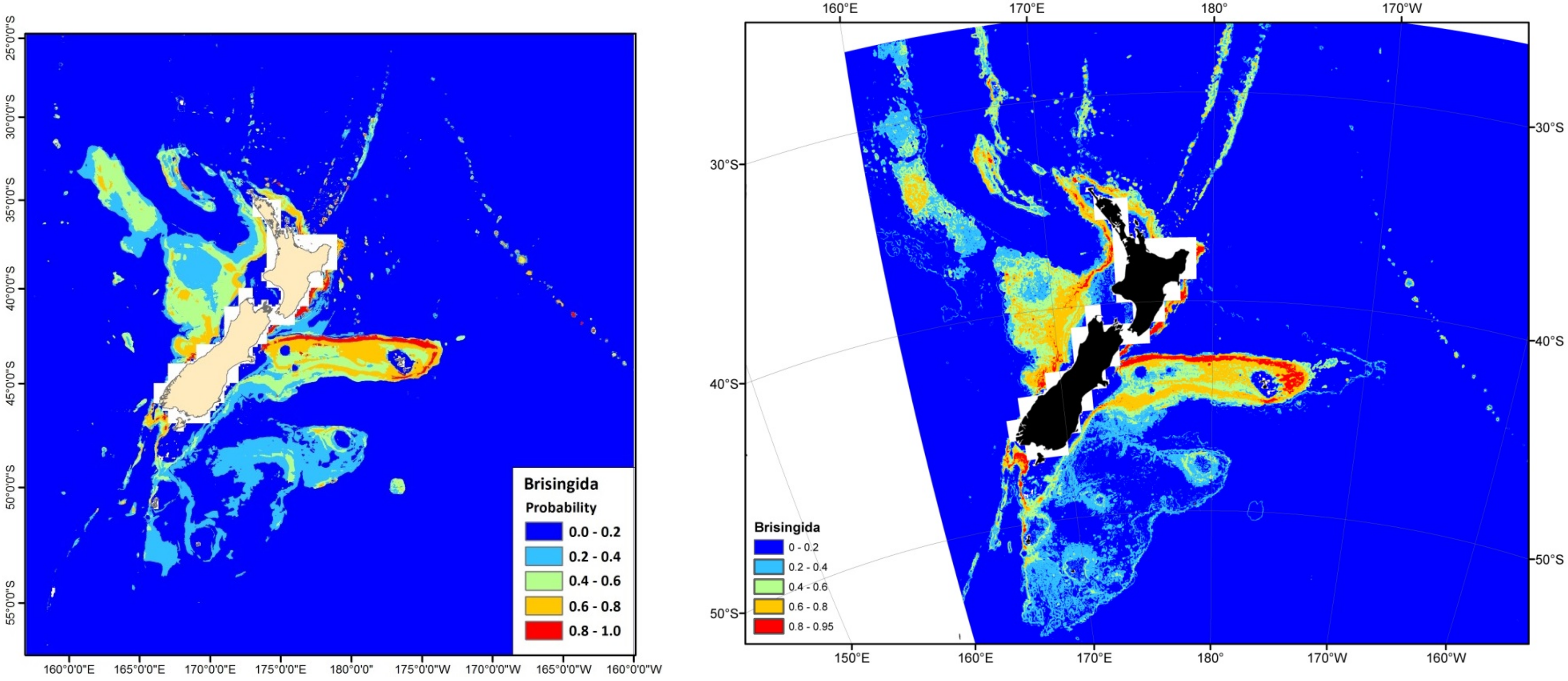


Figure 7: Maxent habitat suitability predictations based on ‘regional’ (left) and ‘global’ (right) environmental data layers for example taxa showing (a) similar predictions, and (b) dissimilar predictions. (key shows probability of habitat suitability from 0 – 1).

(a) Brisingida – similar predictions



(b) Pennatulacea - dissimilar predictions of high habitat suitability

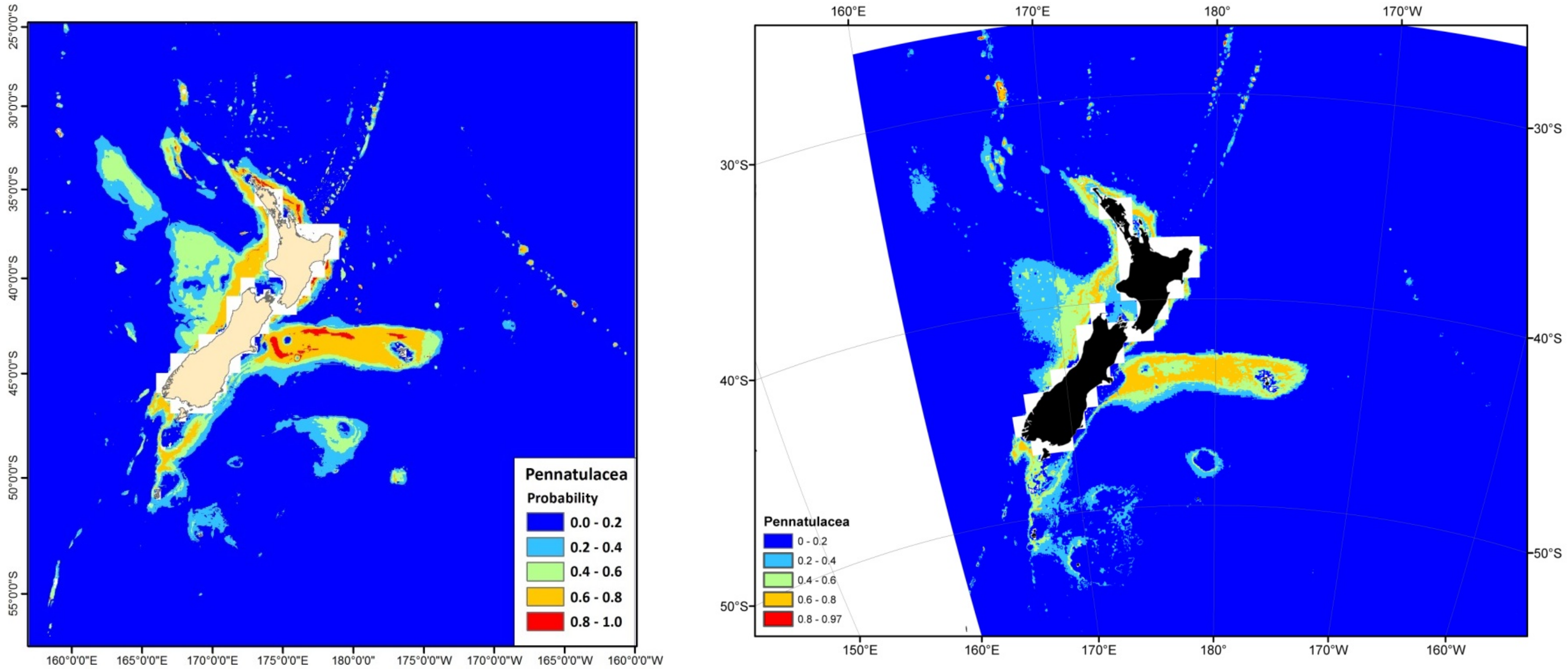


Figure 8: Maxent model predictations based on 'regional' (left) and 'global' (right) environmental data layers for an example taxon (Porifera) showing large areas of the seafloor less than750 m deep as being suitable habitat (more than 0.6 probability). [key shows probability of habitat suitability from 0 – 1].

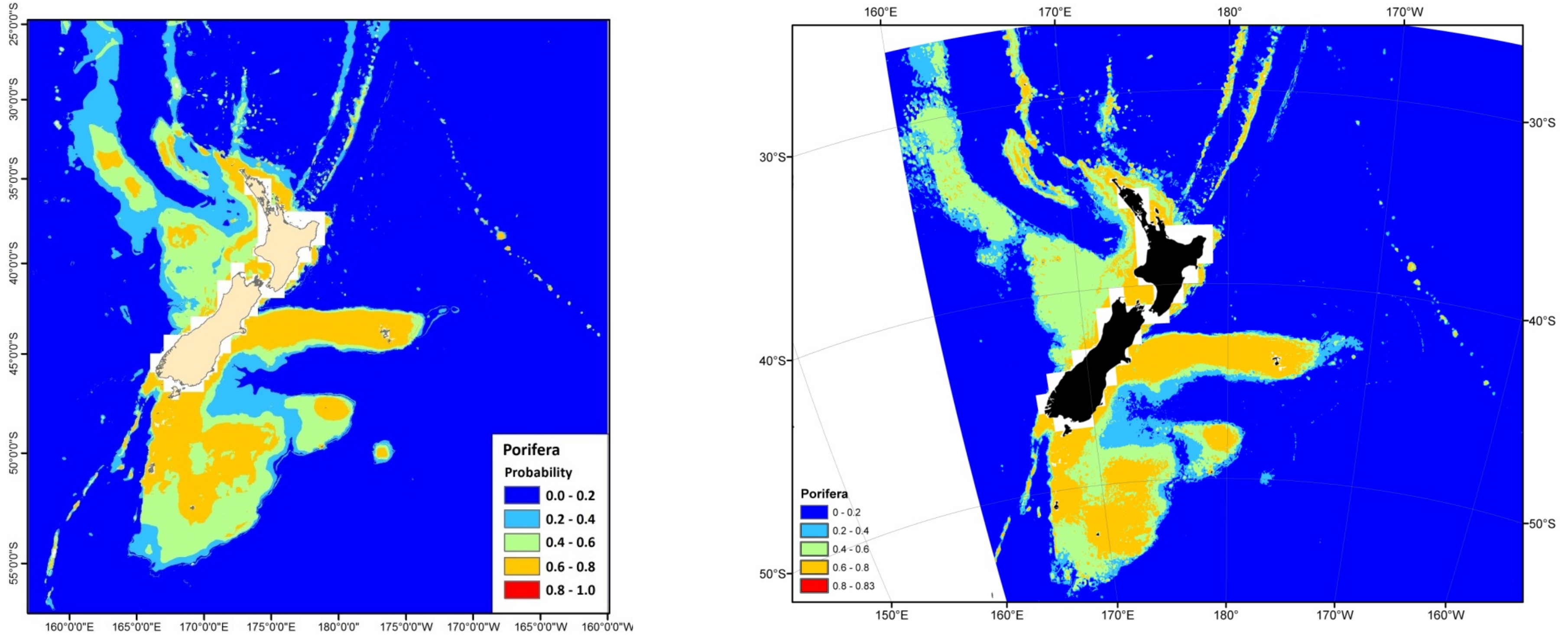
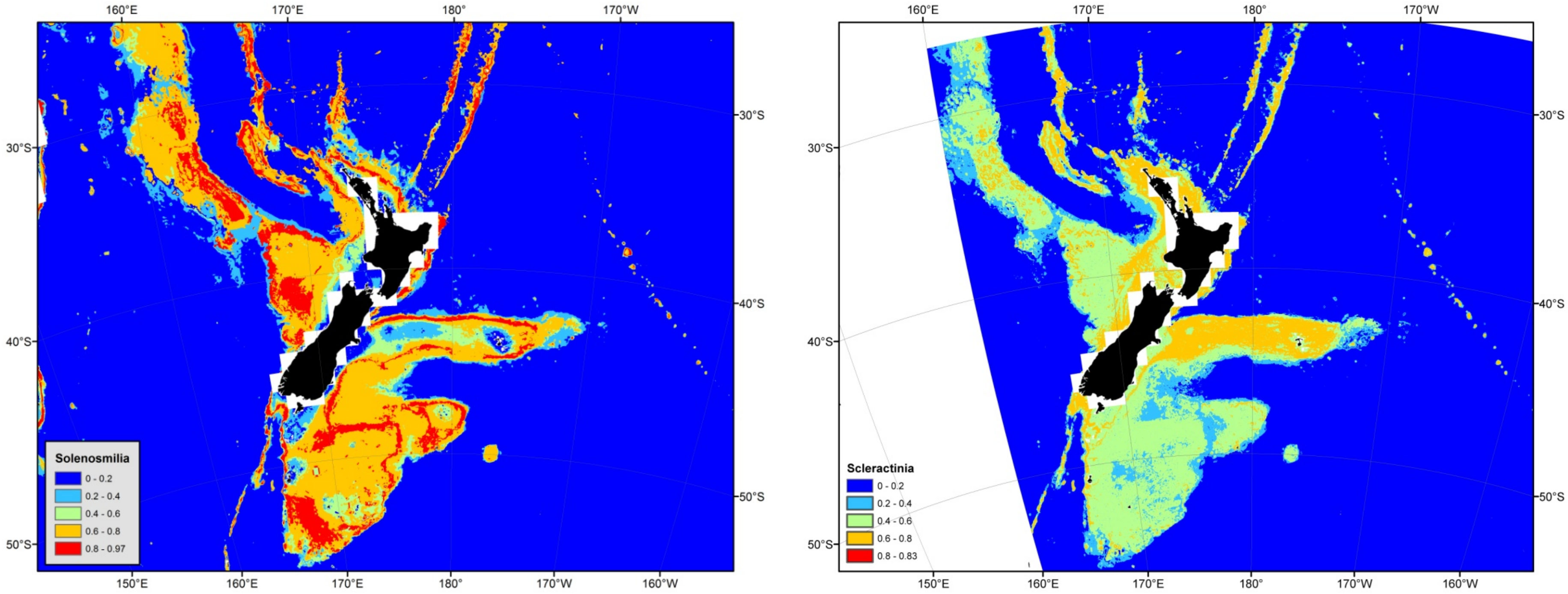


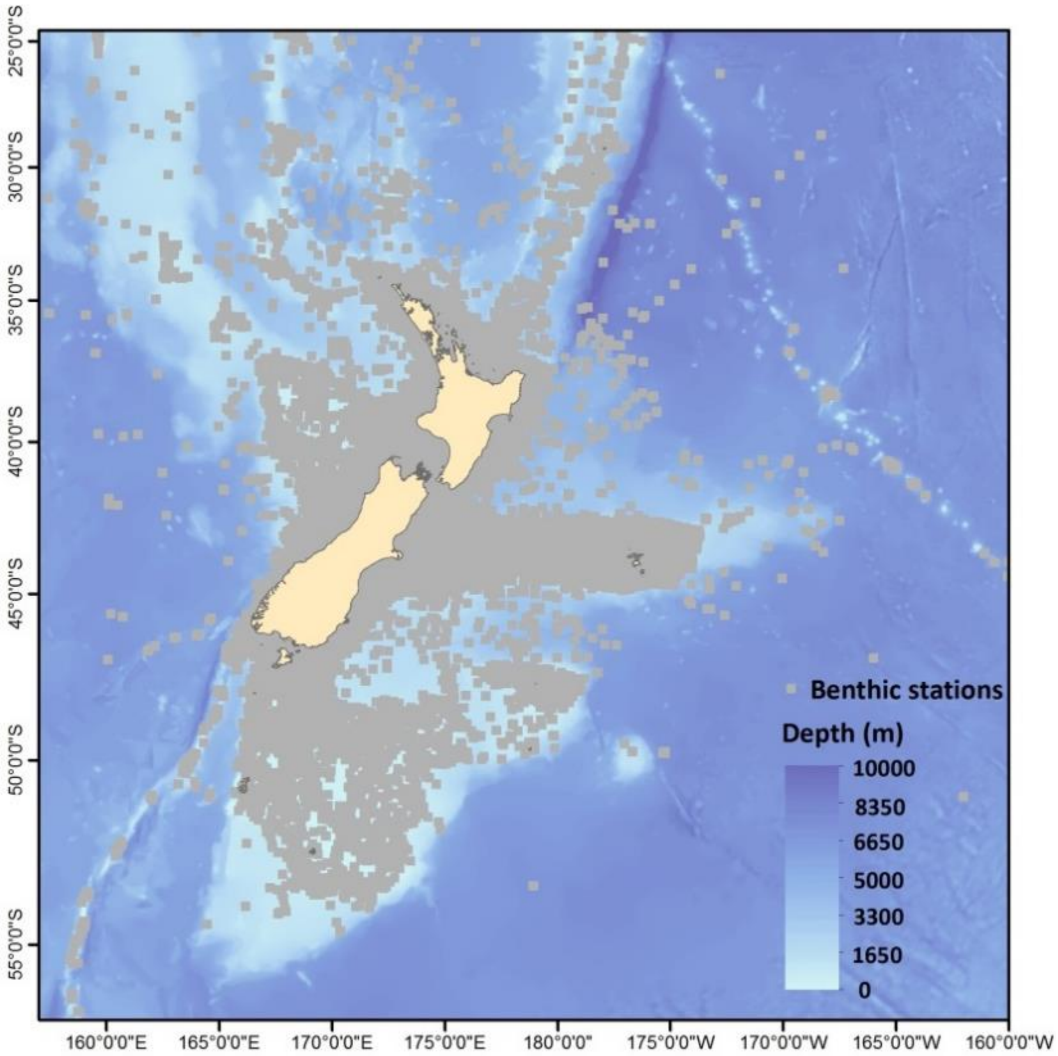
Figure 9: Maxent habitat suitability model predictions for (a) *Solenosmilia variabilis* and (b) the VME indicator taxon Scleractinia using ‘global’ environmental data. [key shows habitat suitability from 0 – 1].



(a) *Solenosmilia variabilis*

(b) Scleractinia

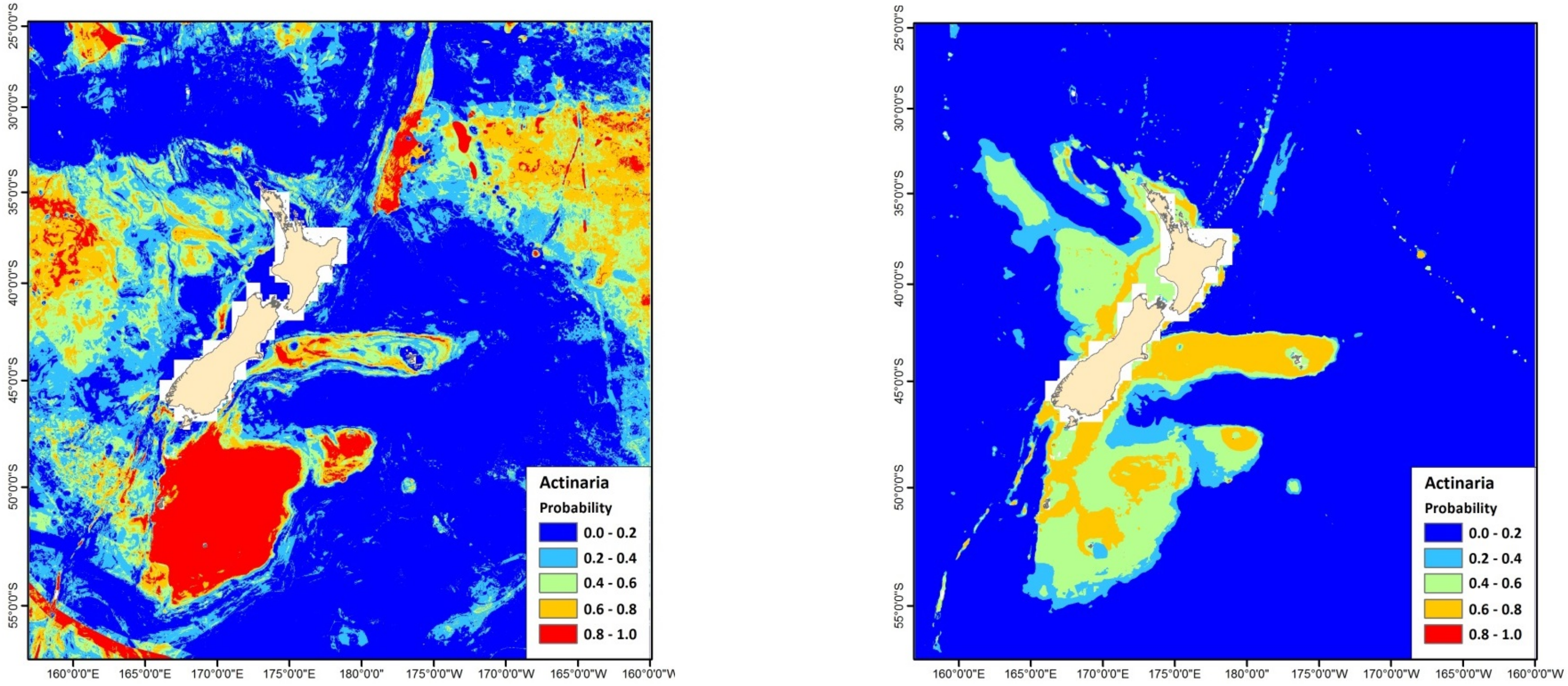
Figure 10: Map showing the distribution of benthic sampling stations in the study area from which records of the ten VME indicator taxa were obtained for the predictive habitat suitability modelling.



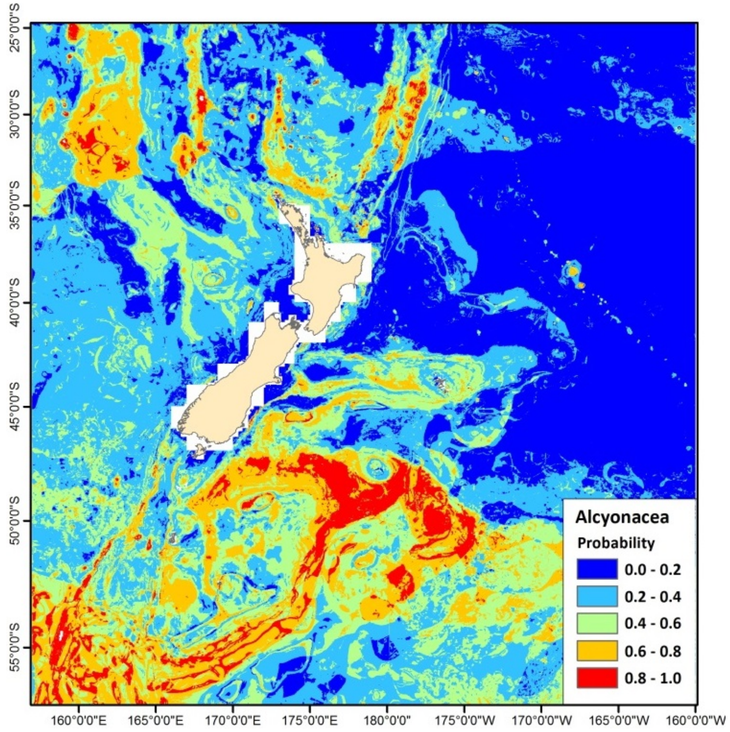
APPENDIX 1

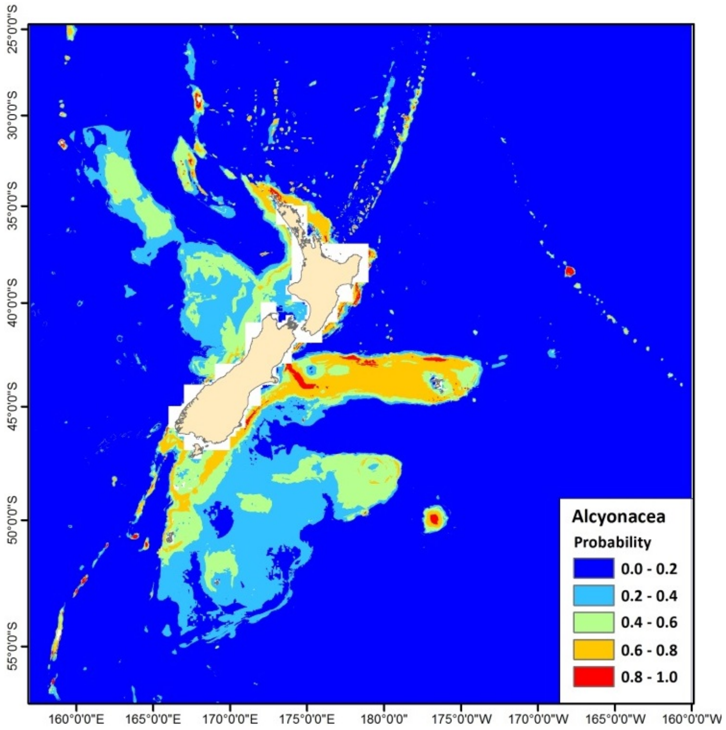
Appendix 1.1: BRT (left) and Maxent (right) predicted habitat suitability for VME indicator taxa using ‘regiona1’ environmental data (key shows probability of habitat suitability from 0 – 1).

(a) Actinaria

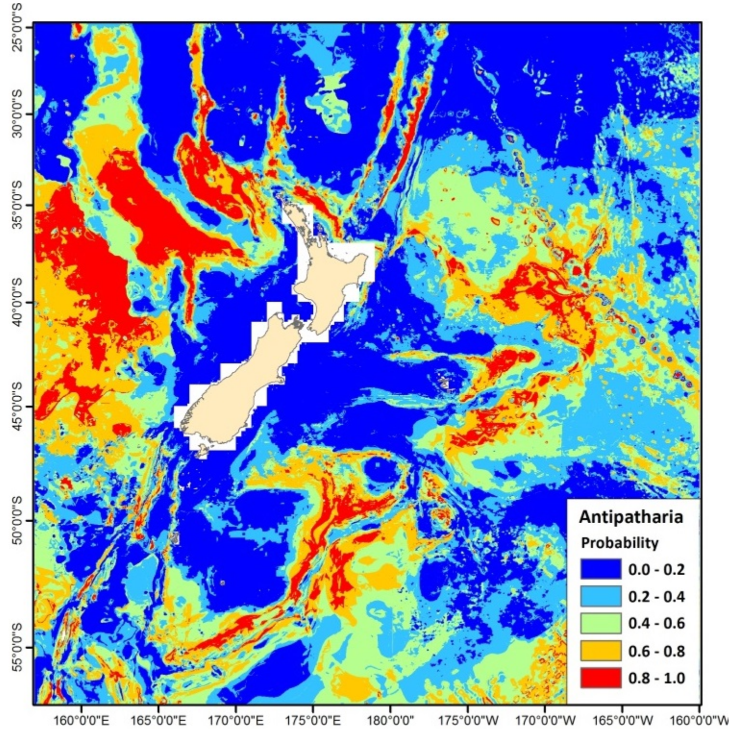


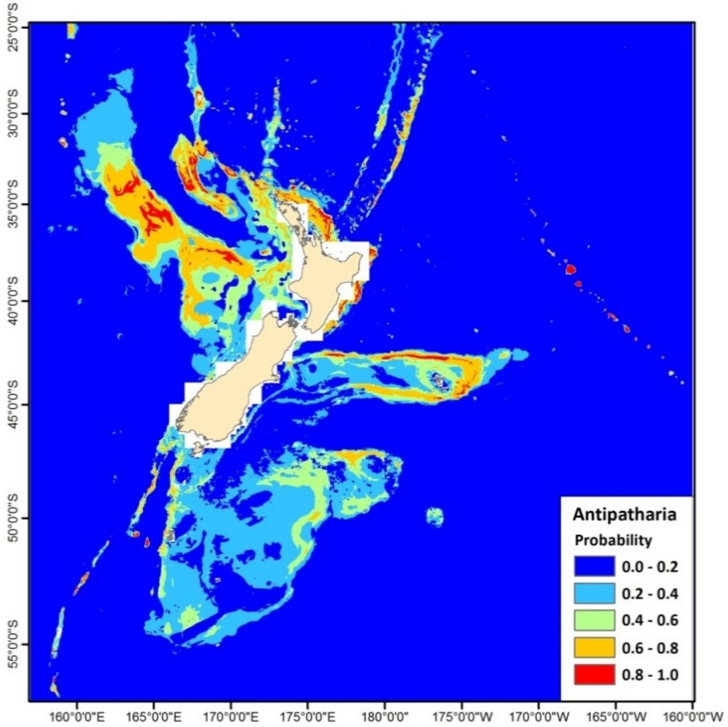
(b) Alyconacea



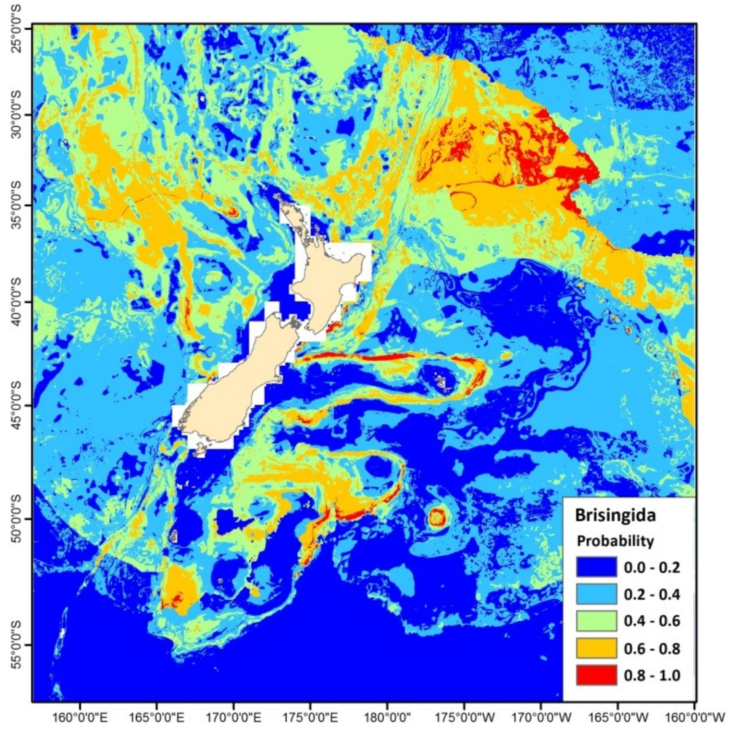


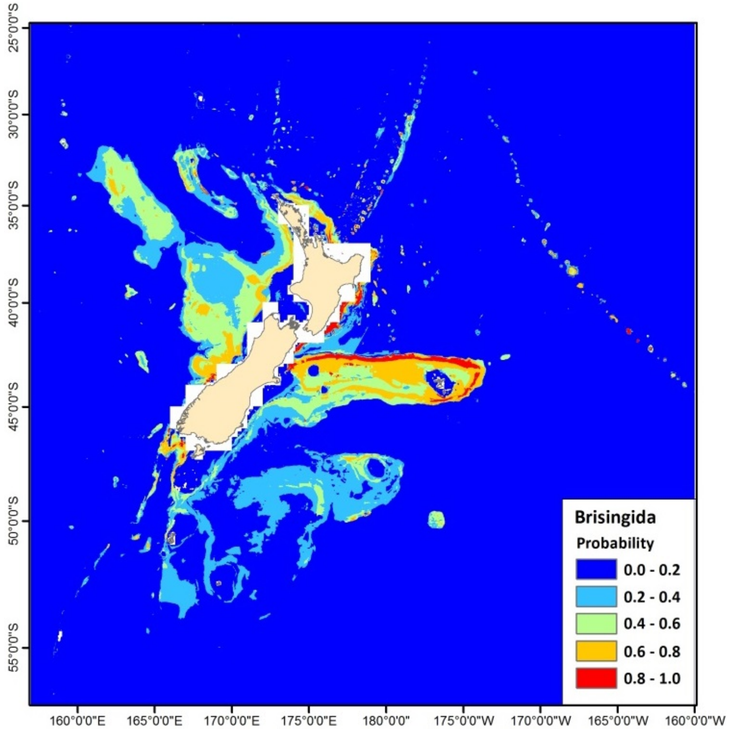
(c) Antipatharia



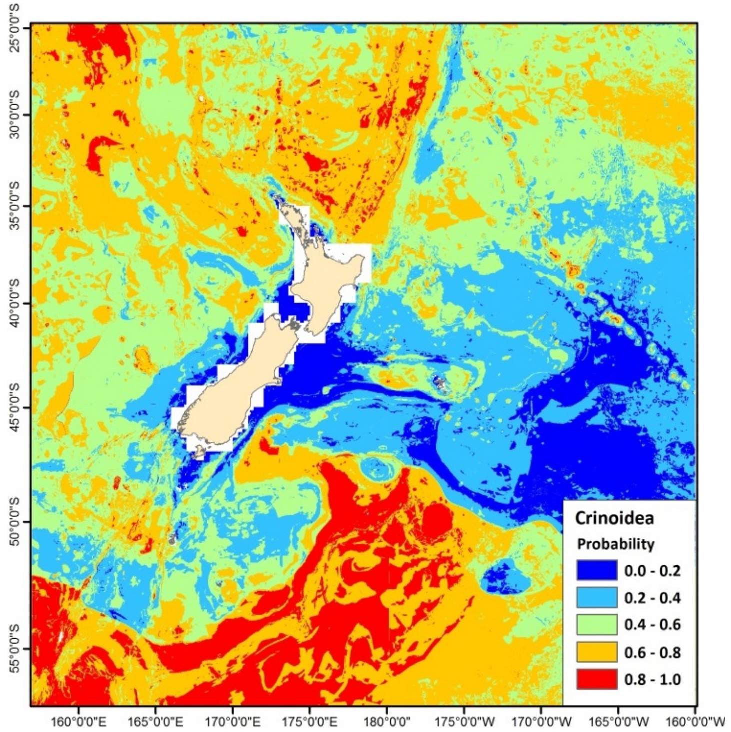


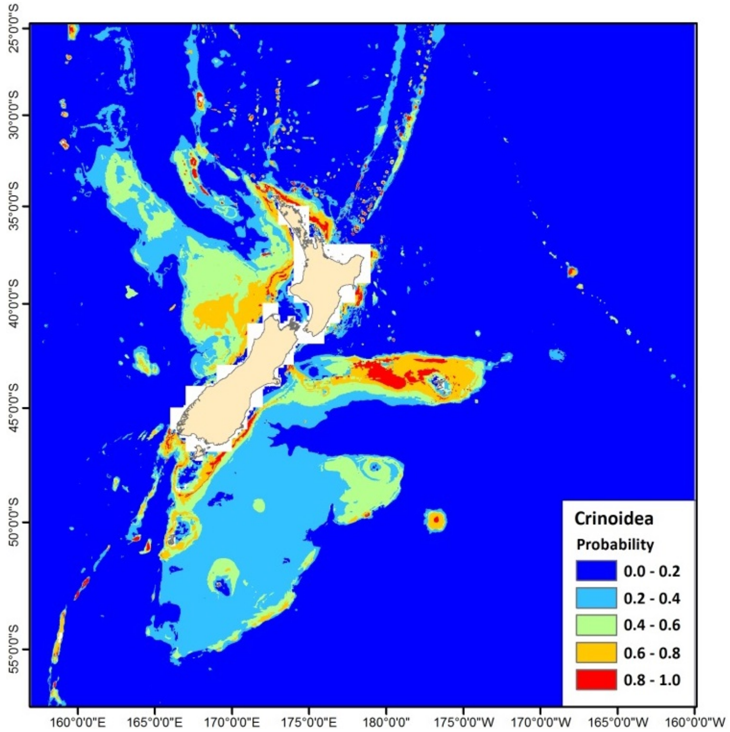
(d) Brisingida



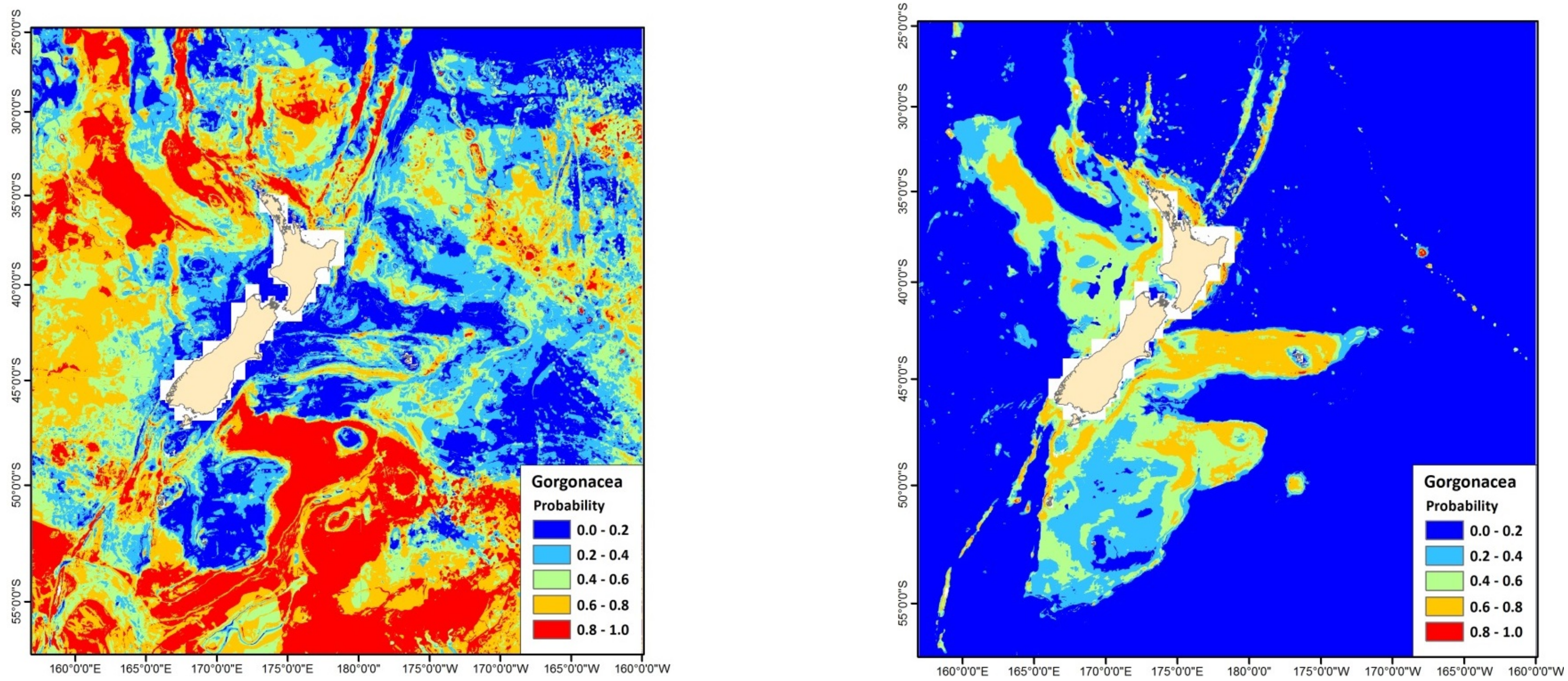


(e) Crinoidea

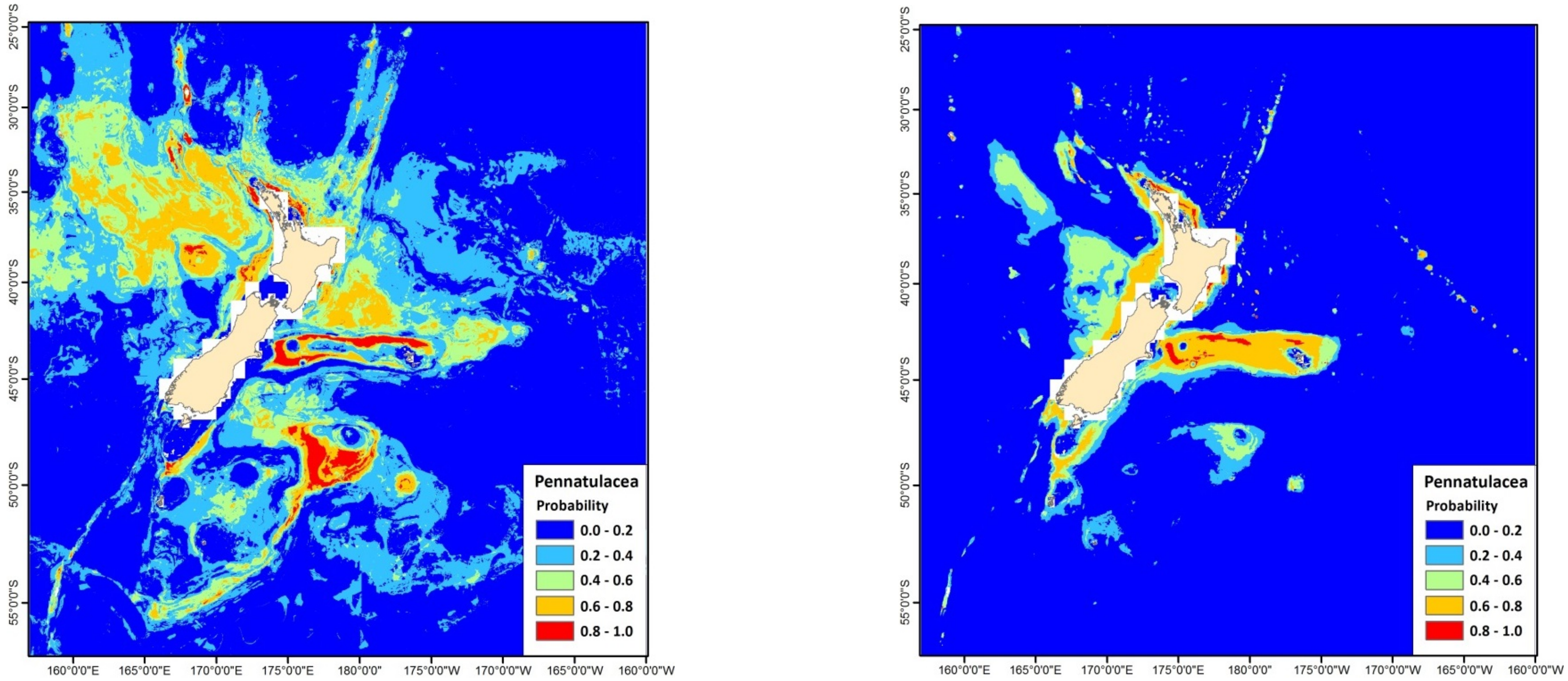

25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W
Crinoidea
Probability
0.0 - 0.2
0.2 - 0.4
0.4 - 0.6
0.6 - 0.8
0.8 - 1.0


25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W
Crinoidea
Probability
0.0 - 0.2
0.2 - 0.4
0.4 - 0.6
0.6 - 0.8
0.8 - 1.0

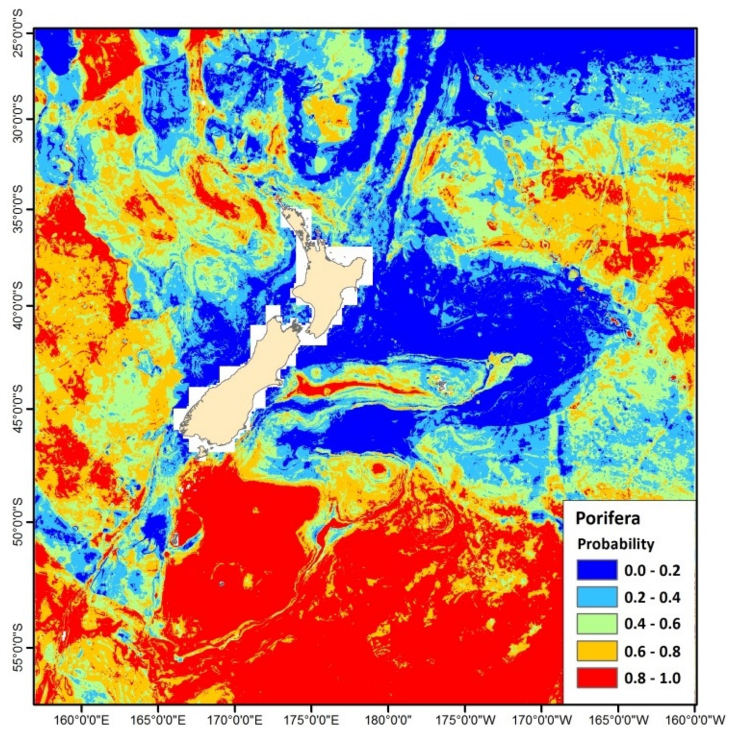
(f) Gorgonacea

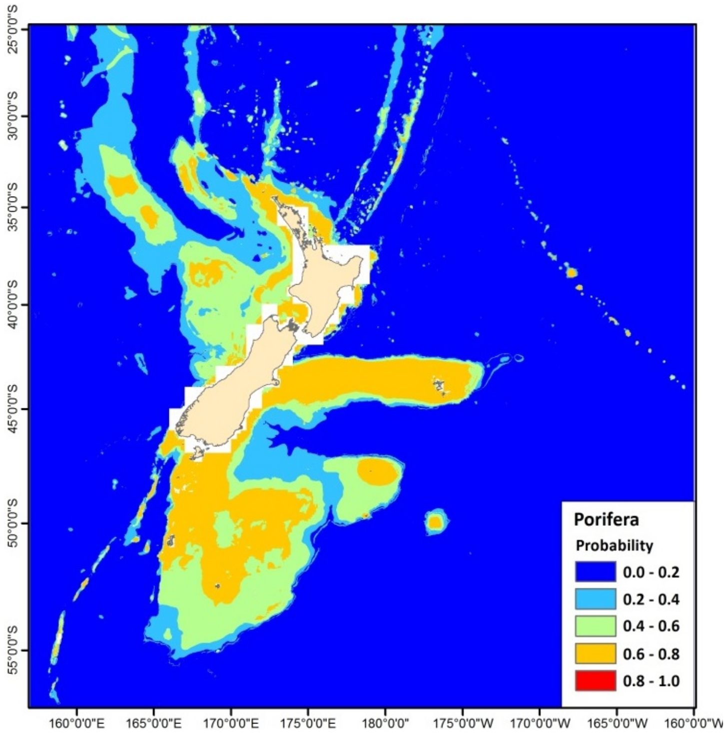


(g) Pennatulacea

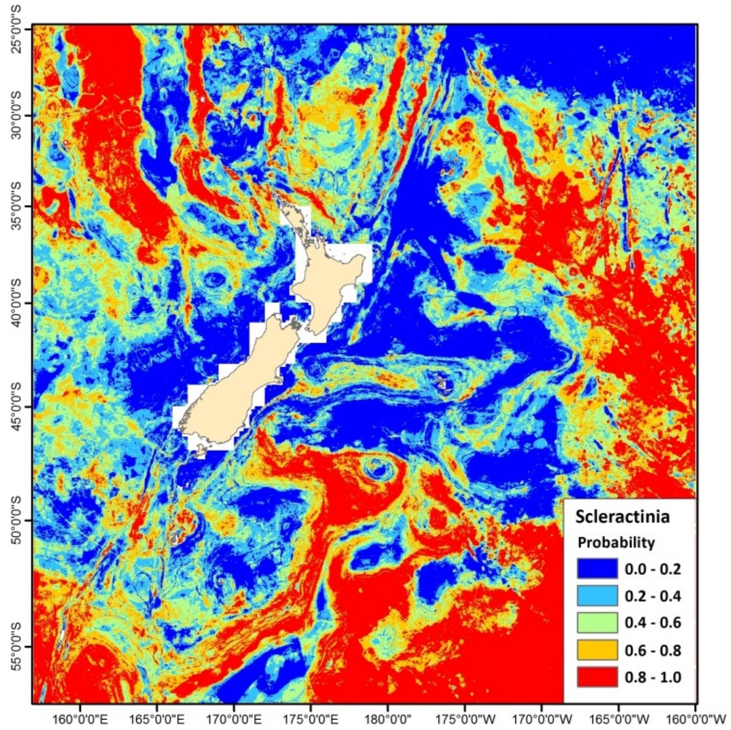


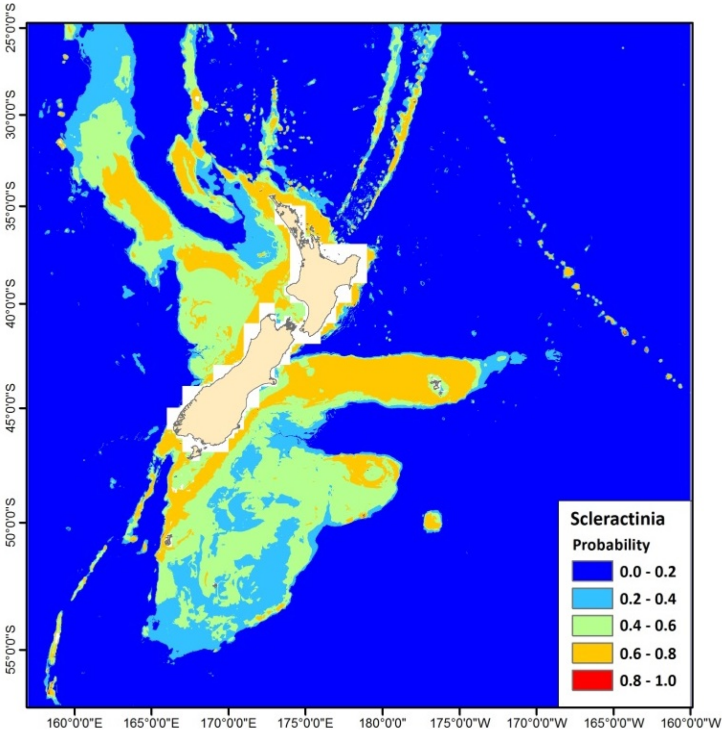
(h) Porifera


Porifera
Probability
0.0 - 0.2
0.2 - 0.4
0.4 - 0.6
0.6 - 0.8
0.8 - 1.0
25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W

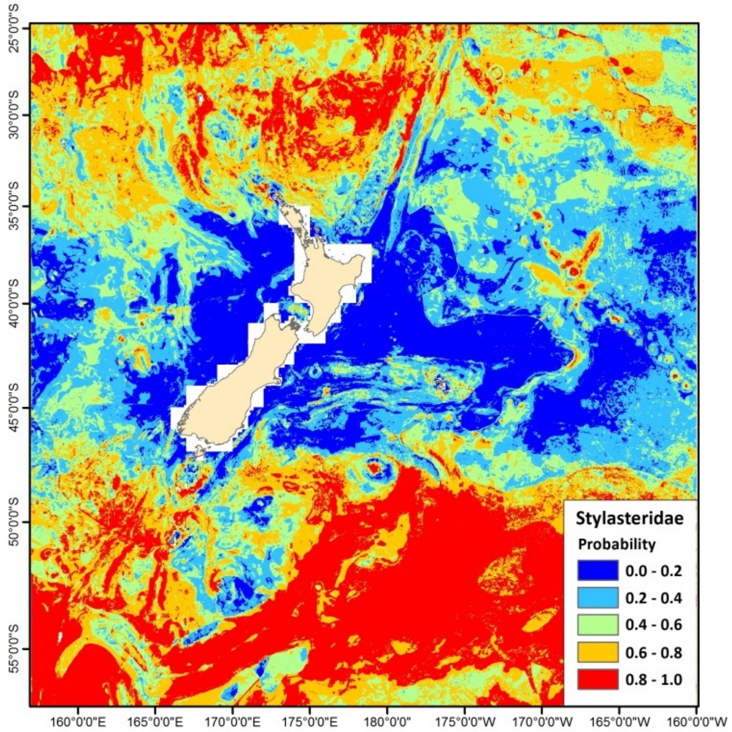

Porifera
Probability
0.0 - 0.2
0.2 - 0.4
0.4 - 0.6
0.6 - 0.8
0.8 - 1.0
25°0'0"S
30°0'0"S
35°0'0"S
40°0'0"S
45°0'0"S
50°0'0"S
55°0'0"S
160°0'0"E
165°0'0"E
170°0'0"E
175°0'0"E
180°0'0"
175°0'0"W
170°0'0"W
165°0'0"W
160°0'0"W

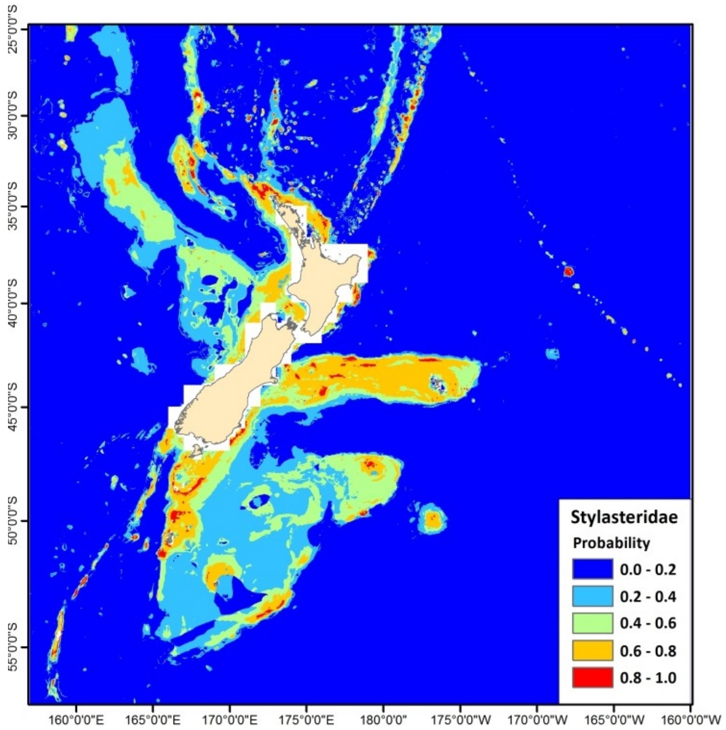
(i) Scleractinia



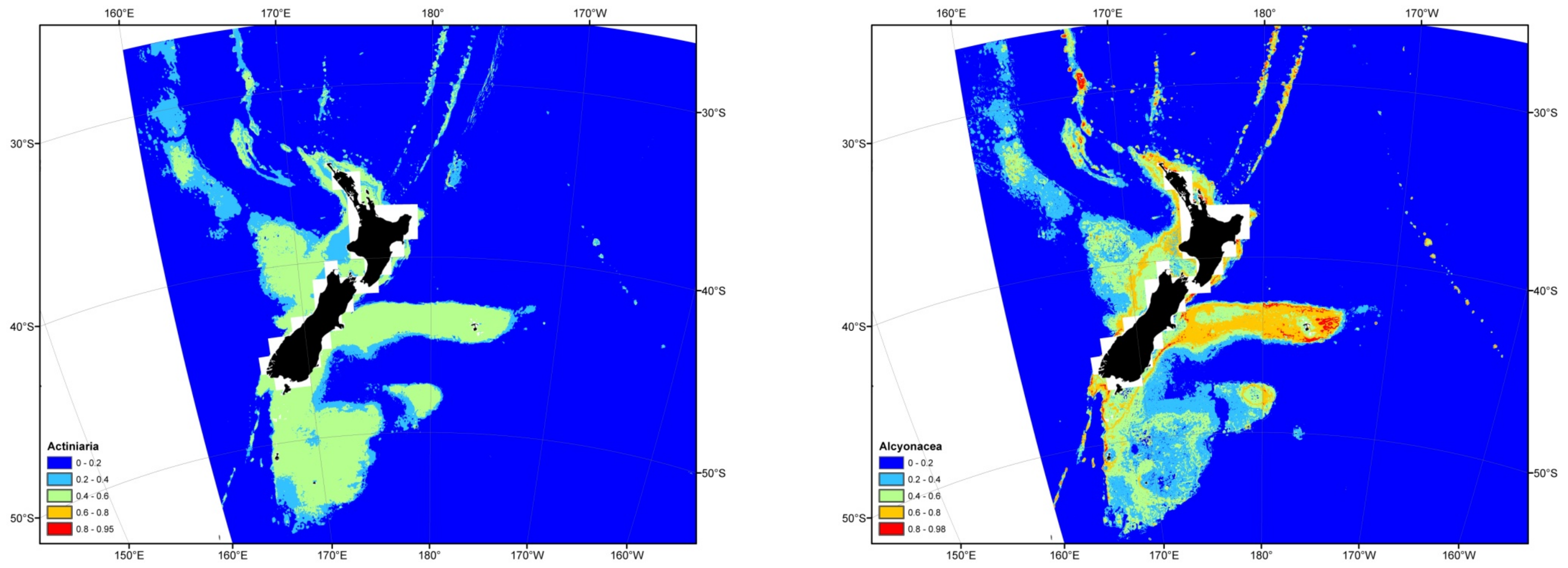


(j) Stylasteridae





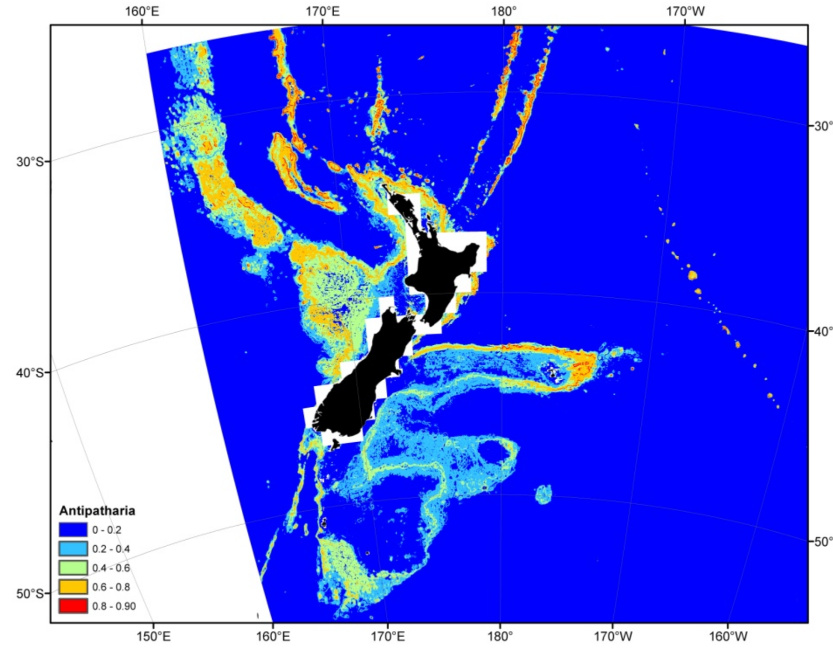
Appendix 1.2: Maxent predicted habitat suitability for VME indicator taxa using ‘global’ environmental data (key shows probability of habitat suitability from 0 – 1).



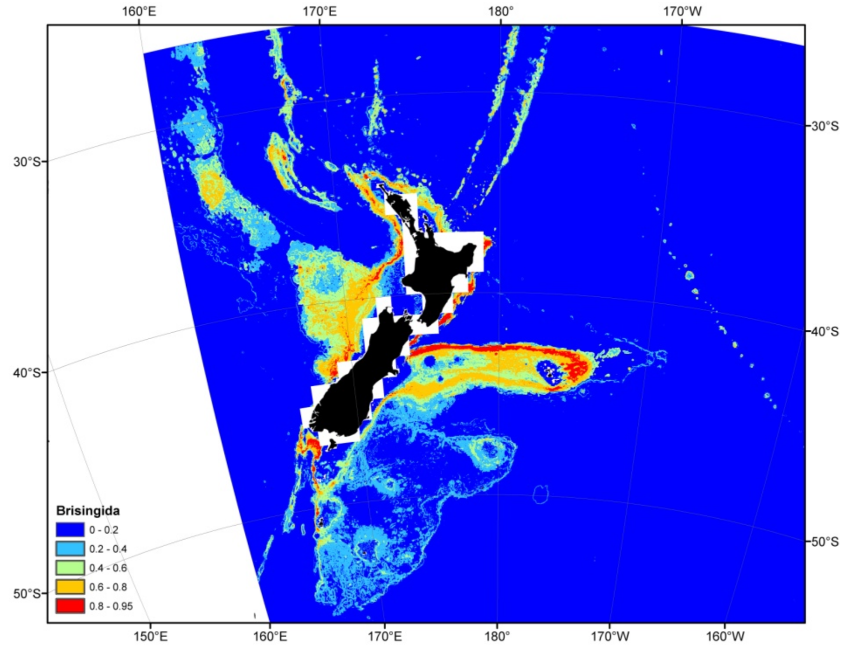
(a) Actinaria

(b) Alyconacea

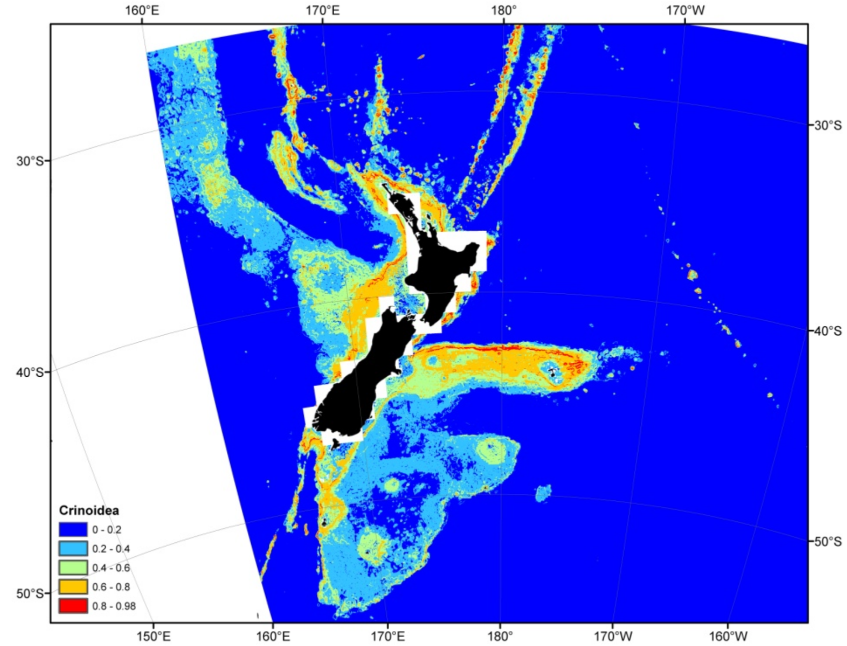
(c) Antipatharia



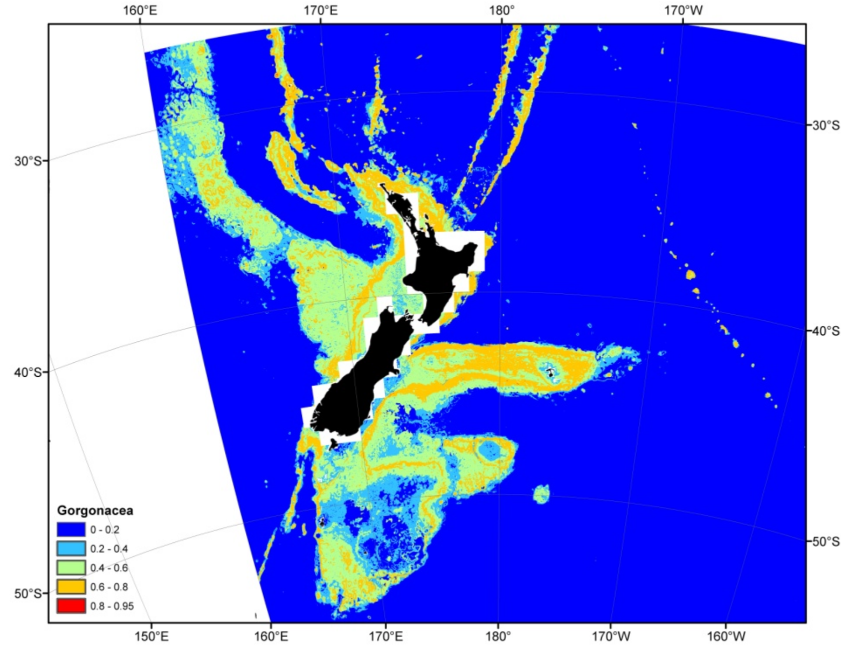
(d) Brisingida



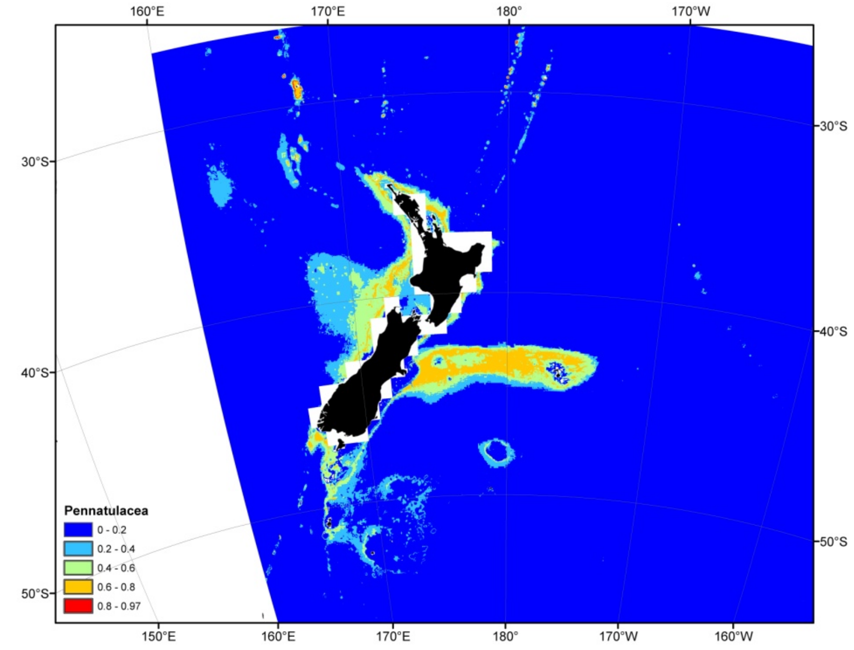
(e) Crinoidea



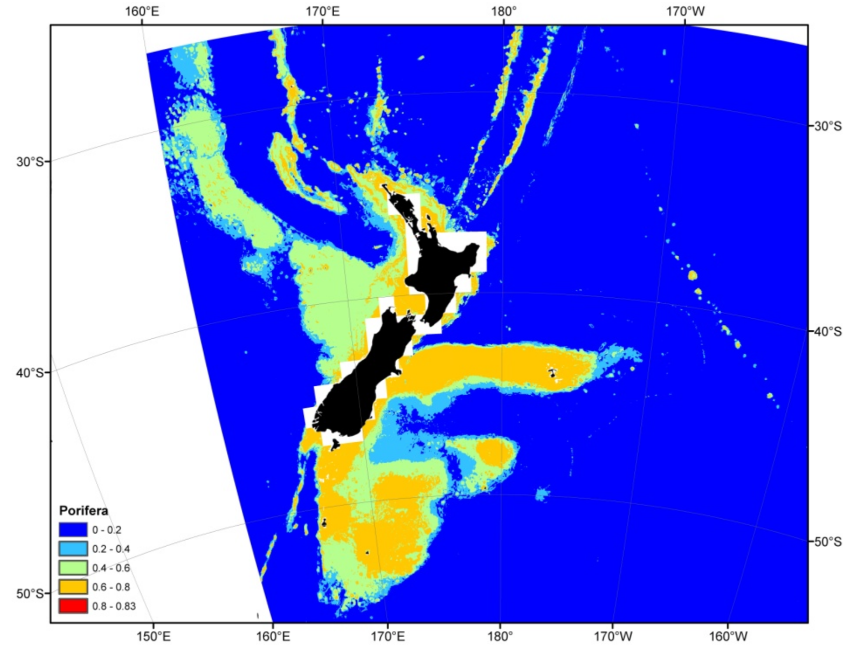
(f) Gorgonacea



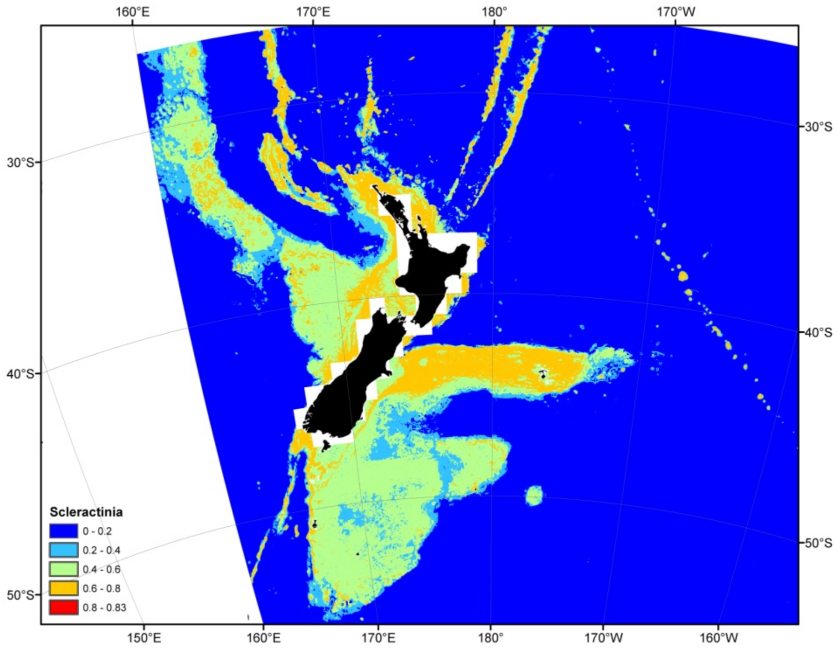
(g) Pennatulacea



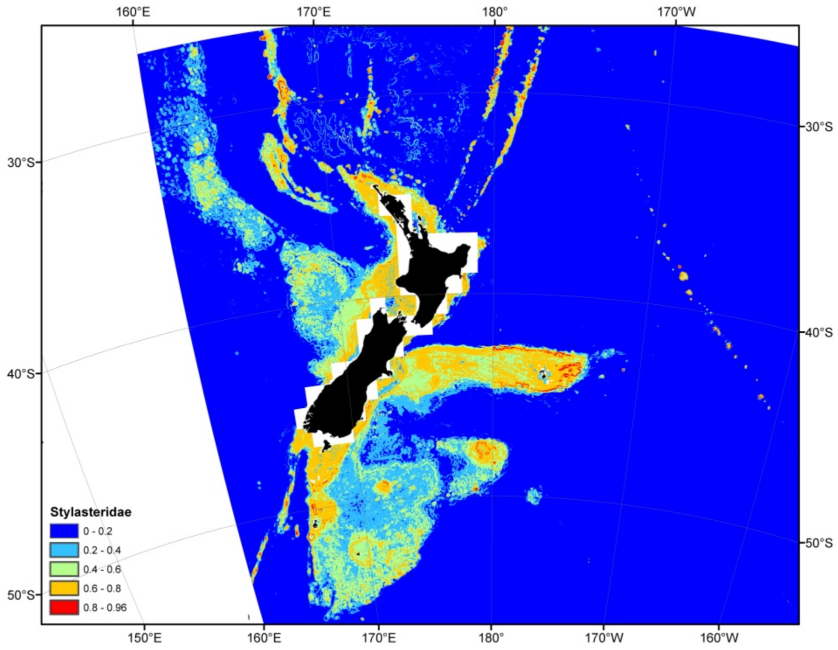
(h) Porifera



(i) Scleractinia



(j) Stylasteridae



APPENDIX 2

Appendix 2.1: The mean percentage contribution of each environmental variable to explaining the habitat suitability of each VME indicator taxon for the Maxent models using 'regional' environmental data.

Variable	Actinaria	Alcyonacea	Antipatharia	Brisingida	Crinoidea	Gorgonacea	Pennatulacea	Porifera	Scleractinia	Stylasteridae
bathy	86.9	72.2	65.4	31.2	61.3	79.5	50.4	86.9	87.1	73.0
botspeed	0.1	0.7	1.3	1.3	0.4	1.3	0.5	0.7	1.2	1.7
disorg	0.8	1.7	2.0	15.1	2.0	0.5	1.7	0.8	0.2	2.2
dynoc	2.5	1.3	7.8	1.0	2.3	3.3	0.4	1.5	0.7	0.5
poc	6.2	5.8	1.4	27.3	4.4	0.5	22.3	1.1	0.3	0.6
slope	0.1	2.3	5.0	7.8	2.1	3.1	0.1	0.6	1.4	4.8
seamount	0.6	8.6	5.4	1.6	16.2	4.0	1.5	0.9	2.2	8.1
sstgrad	0.7	3.9	2.1	1.9	1.6	2.9	3.4	0.2	2.5	3.5
tempbot	1.7	0.6	1.6	2.8	0.8	1.2	0.5	6.9	0.6	2.5
tidalcurr	0.2	0.7	0.5	0.3	2.1	0.7	1.1	0.3	0.1	2.7
vgpm	0.0	2.2	7.5	9.7	6.9	3.0	18.0	0.1	3.6	0.4

Bathy = depth; *botspeed* = bottom speed; *disorg* = dissolved organic matter; *dynoc* = dynamic topography; *poc* = particulate organic carbon flux; *slope* = slope; *seamount* = seamount; *sstgrad* = SST gradient; *tempbot* = bottom temperature; *tidalcurr* = tidal current; *vgpm* = surface water primary productivity.

Appendix 2.2: The mean percentage contribution of each environmental variable to explaining the habitat suitability of each VME indicator taxa for the five simplified BRT models using 'regional' environmental data.

Variable	Actinaria	Alcyonacea	Antipatharia	Brisingida	Crinoidea	Gorgonacea	Pennatulacea	Porifera	Scleractinia	Stylasteridae
bathy	4.81	5.00	6.03	32.43	5.37	5.59	34.61	5.13	8.27	4.24
botspeed	4.31	10.48	3.35	6.92	2.48	6.81	4.02	5.78	10.97	11.07
disorg	3.99	17.22	14.05	10.70	12.52	12.11	3.33	32.75	12.22	11.09
dynoc	9.54	18.03	16.31	8.61	18.70	11.37	13.31	7.86	13.97	11.89
poc	5.57	4.40	35.39	3.28	4.39	26.56	6.27	6.52	16.63	11.98
slope	4.66	6.96	7.56	4.67	7.94	5.78	4.27	3.54	6.97	4.40
seamount	0.10	3.86	1.08	0.05	12.69	1.25	0.00	1.76	0.76	6.15
sstgrad	3.56	8.63	4.28	4.08	7.67	6.91	7.36	8.32	8.37	6.56
tempbot	34.67	9.89	3.10	18.53	6.97	4.69	8.05	4.83	5.34	4.43
tidcurr	3.55	4.96	2.12	7.27	4.98	4.99	6.78	13.42	7.15	8.99
vgpm	25.24	10.58	6.74	3.47	16.28	13.93	12.00	10.09	9.34	19.20

Bathy = depth; *botspeed* = bottom speed; *disorg* = dissolved organic matter; *dynoc* = dynamic topography; *poc* = particulate organic carbon flux; *slope* = slope; *seamount* = seamount; *sstgrad* = SST gradient; *tempbot* = bottom temperature; *tidalcurr* = tidal current; *vgpm* = surface water primary productivity.

Appendix 2.3: The mean percentage contribution of each environmental variable to explaining the habitat suitability of each VME indicator taxa for Maxent models using ‘global’ environmental data.

Variable	Actinaria	Alcyonacea	Antipatharia	Brisingida	Crinoidea	Gorgonacea	Pennatulacea	Porifera	Scleractinia	Stylasteridae
Depth	14.5	20.0	12.0	40.5	56.4	25.4	37.9	13.3	2.8	14.3
POC	6.7	9.4	1.2	37.9	14.3	0.7	34.8	0.9	1.0	2.8
Slope	0.2	4.7	7.2	2.3	5.7	3.9	0.3	0.6	2.2	7.5
Temperature	77.2	40.2	53.5	6.2	18.4	48.9	10.0	69.6	55.5	49.7
Dissolved Oxygen	1.2	3.3	9.3	6.8	2.2	4.1	3.4	1.0	1.9	4.1
Salinity	0.3	2.8	0.6	6.2	3.0	0.9	3.0	1.5	0.4	2.8
Calcite	-	19.6	16.2	-	-	16.1	-		-	-
Aragonite	-	-	-	-	-	-	10.6		36.2	18.7
Silica	-	-	-	-	-	-	-	12.9	-	-