Updated habitat suitability modelling for protected corals in New Zealand waters

DRAFT REPORT

Prepared for Department of Conservation

October 2019

Owen Anderson, Fabrice Stephenson, Erik Behrens

NIWA

Contents

Tables

Figures

Executive summary

Many protected coral species occur as bycatch in commercial fisheries in the New Zealand region. To determine the overlap between commercial fishing and corals, it is first necessary to determine and predict the spatial extent of corals. This work expands on previous protected coral habitat suitability modelling studies by utilising updated modelling techniques, incorporating additional coral presence records from recent commercial and research sampling surveys, and by using regional environmental predictor layers for the current and future climate conditions based on NIWA's Earth System Model (ESM).

We applied sophisticated habitat suitability modelling techniques that consider spatial autocorrelation in the sampling data, estimate precision of the predicted distributions, combine two model types (boosted regression trees and random forests), and assess model performance.

Models were produced for all protected coral taxa considered in initial consultations with stakeholders. These comprised four separate reef-forming scleractinian coral species, *Enallopsammia rostrata*, *Solenosmilia variabilis*, *Goniocorella dumosa*, and *Madrepora oculata*; the gorgonian octocoral genera *Paragorgia* (bubblegum corals), *Primnoa* (primnoid seafans), *Corallium* (precious corals); and *Keratoisis* & *Lepidisis* (bamboo corals) combined; two antipatharian (black) coral genera, *Bathypathes* and *Leiopathes*; and two genera of stylasterid hydrocorals, *Errina* and *Stylaster*. This selection was based on the need to produce models that cover a range of the protected coral taxa for which there is also a relatively large number of records.

Environmental predictors were derived primarily from outputs of the New Zealand Earth System Model but several fixed predictors, including revised and updated sediment data layers, seafloor slope, and seamount (as a categorical variable) were also considered.

Model coefficients were used to produce two sets of prediction grids for each model type; one for present-day environmental conditions (means from the period 1995 to 2014), and one for the predicted environmental conditions at the end of the 21st century (means from the period 2080 to 2099), assuming only moderate mitigation and adaptation to climate change.

Model performance was acceptable for all taxa, and although predicted species distributions largely agreed with previous studies, additional presence records extended the predicted distributions into new areas for some taxa. Predicted habitat suitability at the end of the century tended to flatten out for several taxa, with lowered suitability in areas of currently high suitability and vice versa. For the region as a whole, future habitat suitability ranged from somewhat less suitable (e.g. *Corallium*) to somewhat more suitable overall (e.g. *Enallopsammia rostrata*), across the 12 taxa examined. For some taxa, especially the hydrocorals, the predicted future habitat suitability remained largely unchanged due to the importance of temporally fixed parameters in the models.

The risk to corals from interaction with fishing gear was assessed by comparing predicted coral distributions with the aggregated swept area from recent bottom fishing for inshore and deepwater fisheries combined. Overlaying the regions of greatest habitat suitability with the regions most highly fished revealed considerable variability in vulnerability among taxa, both in degree and location. The greatest overlaps were seen for hydrocorals and the shallower scleractinian species, whereas the deeper scleractinians, gorgonians, and black corals were less vulnerable.

1 Introduction

All coral species in the Orders Antipatharia and Scleractinia, gorgonian octocorals in the Order Alcyonacea, and stylasterid hydrocorals in the Order Anthoathecata, are protected under the Wildlife Act (1953) and a later (2010) amendment. Many of these protected coral species occur as bycatch in commercial fisheries in the New Zealand region (e.g., Anderson et al. 2017).

In order to refine our understanding of the overlap between commercial fishing effort and corals, and to assess potential fishing impacts across their distribution, it is necessary to quantify and predict the spatial extent of corals in relation to these impacts. This project expands on the work undertaken by Anderson et al. (2014), by carrying out improved and refined habitat suitability modelling to predict species distributions using new data. These new data include coral records collected by researchers and the Department of Conservation's (DOC) Conservation Services Programme (CSP) Observer programme during the past four years and identified by specialists (Tracey et al. 2017), coral records from shallow depths (less than 200 m), and regional environmental predictor layers for the current and future climate conditions based on the New Zealand Earth System Model (NZESM).

Comparison of the predicted distributions of coral taxa with current fishing effort is made using bottom trawl footprint data compiled for recent fishing years (Baird & Mules 2019). Coral data from less than 200 m was included in the models where available, and the predictions cover the shallowest depths possible within the limits of the available environmental predictor layers.

Updating the predicted distribution maps for protected corals enables improved definition of suitable current and future habitat, helps to assess risk from commercial fishing, and informs the management of these fragile and long-lived fauna.

The specific objectives of this project (DOC19301/POP2018-01) are:

1. To carry out improved habitat suitability modelling for protected corals in the New Zealand region.
2. To help identify areas of risk from interactions with commercial fishing gear.

2 Methods

Various techniques exist for predicting species distributions through the spatial estimation of habitat suitability, and there has been a substantial development of these in recent years, taking advantage of increased computing power, machine learning algorithms, availability of global and regional datasets of environmental variables, the ongoing sampling of the world's oceans, and accessible species record databases.

The most commonly used of the available habitat suitability modelling methods include:

- Generalised Linear Models (GLMs and GAMs) (McCullagh & Nelder 1989, Hastie & Tibshirani 1990)
- Maximum Entropy (Maxent) (Phillips et al. 2006)
- Random Forests (RF) (Brieman 2001)
- Boosted Regression Trees (BRT) (Elith et al. 2008)
- Genetic Algorithm for Rule-Set Production (GARP) (Stockwell 1999)
- Multivariate Adaptive Regression Splines (MARS) (Friedman 1991)

- Ecological Niche Factor Analysis (ENFA) (Hirzel et al. 2002)
- Artificial Neural Networks (ANNs)
- BIOCLIM (Nix 1986)

Although the approaches and underlying structures differ, each of these models essentially explores the relationship between point-sampled species occurrence records (and usually absence or background records), and spatially continuous environmental variables. These models predict the likelihood of occurrence of (or habitat suitability for) a species, or group of species, across unsampled environmental space (Reiss et al. 2015, Vierod et al. 2014). For this project we have chosen two techniques that are able to account for complex non-linear correlations between species occurrence and environmental predictors, and have proven to work well with New Zealand data (e.g., Georgian et al 2019, Bowden et al. in press) – Boosted Regression Trees and Random Forests.

2.1 Assembly of coral presence and absence data

A database containing records of benthic invertebrate samples from the South Pacific region is regularly updated by NIWA staff for use in various marine biodiversity research projects, including HSM studies. The database includes taxa which are considered Vulnerable Marine Ecosystem (VME) indicator taxa (e.g., various coral groups and sponges) and taxa that are considered indicators of VME habitat (i.e., those strongly associated with VMEs; crinoids and brisingid starfish) (Parker et al. 2009). As such the database records include all protected New Zealand coral species.

Coral presence data used in the various previous analyses of protected coral species distributions was then augmented by several years of additional records. These data have come from fisheries observer sample collections, research surveys, and from records from overseas museums and research institutes. These revised and updated datasets of coral records form the basis of this new study. In addition, we have included coral presence data from shallow waters (less than 200 m) to enable the prediction of coral distributions to the shallowest depths possible within the limits of the environmental predictor layers. Coral records were recorded at different taxonomic levels in the database (high taxonomic resolution often was not possible for records based on non-retained specimens); the focus of the current study was to produce models at the genus and species level where possible.

The coral occurrence dataset is a subset of a larger database of position records comprising all research survey stations at which all organisms in the sample were identified, including stations with no corals. This dataset, comprising over 60 000 records within the New Zealand EEZ, was used to provide the absence location data for earlier HSM models, and was used again in this study, expanded as necessary to include data points from stations shallower than 200 m. For each taxon, a set of absence data was randomly selected from the coral occurrence dataset, with the number of records equal to that of the presence data (following the recommendations of Barbet-Massin et al. 2012); the presence and absence records were then combined to provide the basis for model construction.

2.2 Selection of taxa to model

The set of taxa modelled in the two previous related studies (Anderson et al 2014, Anderson et al. 2015) were used as a starting point for taxa selection in the current study. Selection was initially guided by the coral species listed in the DOC Threatened Species List (Freeman et al. 2010) and a DOC marine invertebrate expert panel list (Freeman et al. 2013) (Table 3 1). Further details of selection criteria and rationale for species groupings can be found in Anderson et al. (2014).

Although models combining groups of taxa may be less reliable due to the conflicting effects of the differing environmental tolerances of individual species within them, this issue needed to be balanced against the lack of resources available to the current project to produce models for large numbers of individual species, and the limited number of individual species with sufficient presence data to produce robust models. The available presence data were assessed to consider additions of records of protected coral species collected since the most recent database update.

Limited resources and the greater complexity now used in the analyses restricted the final number of taxa modelled to the twelve shown in Table 2-1, one less than the thirteen taxa in the previous study. This complexity – combining multiple independent models, calculating cross-validated model performance statistics, resampling for precision estimation, dealing with spatial auto-correlation – now represent NIWAs current standard approach to producing habitat suitability maps for VME indicator taxa in the New Zealand and wider Pacific region (e.g., Anderson et al. 2016b, Rowden et al. 2017, Georgian et al. 2019), and has been adopted internationally. Therefore, we assigned a priority ranking to each candidate taxon (Table 2-1).

Table 2-1: Protected coral taxa modelled. Brief description, depth range of records, number of recorded observations available for the models (number available in the previous study, Anderson et al. (2015), in parentheses), and a priority rank. –, not modelled previously.

Order	Taxon	Description	Depth range (m)	N. records	Priority
Scleractinia	*Enallopsammia rostrata*	Reef-forming coral	186–2620	307 (130)	1
	Solenosmilia variabilis	Reef-forming coral	130–2620	472 (311)	1
	Goniocorella dumosa	Reef-forming coral	94–1595	699 (212)	1
	Madrepora oculata	Reef-forming coral	89–2882	251 (126)	1
Alcyonacea	*Paragorgia* spp.	Bubblegum coral (tree-like)	152–2161	221 (98)	2
	Primnoa spp.	Primnoid sea-fans (tree-like)	150–1611	124 (73)	2
	Corallium spp.	Precious coral	108–2427	99 (–)	2
	Keratoisis & *Lepidisis* spp. combined	Bamboo corals (tree-like)	138–2842	565 (241)	1
Antipatharia	*Bathypathes* spp.	Black coral (tree-like)	161–1831	203 (75)	1
	Leiopathes spp.	Black coral (tree-like)	110–1657	199 (67)	3
Anthoathecata	*Errina* spp.	Hydrocorals (small, hard)	66–2771	246 (–)	2
	Stylaster spp.	Hydrocorals (small, hard)	96–2094	225 (–)	3

2.3 Environmental predictors

A wide range of environmental parameters that can potentially have an impact on seafloor-dwelling invertebrates are available as outputs from ESM models, representing both current condition and predicted conditions at some time in the future (e.g., 2100).

The recently developed New Zealand Earth System Model (NZESM) incorporates component models of ocean biogeochemistry and other aspects of biology and chemistry that provide a highly complex model of the climate system (Behrens et al. 2019).

The model differs from the UK Earth System Model essentially in two respects: firstly, it includes a high-resolution regional ocean model for the seas around New Zealand. Secondly it includes a representation of the variability of solar radiation in atmospheric chemistry. Both these changes manifestly affect climate of the Southern Hemisphere. A preliminary evaluation shows that a cold sea surface temperature bias found in the UKESM in the New Zealand region is improved in the NZESM. Solar input drives climate variability particularly at southern high latitudes (such that when the Sun is more active, pressure over Antarctica tends to be lower than otherwise). The NZESM reproduces this association.

This NZESM can produce projections out to 200 years into the future (Williams et al. 2016), but for this analysis we focussed on predicted environmental conditions for an historical reference period (1995–2014) and for the end of the current century (2080–2099). The predicted future conditions were based on the SSP2 (4.5 W/m^2) pathway, which describes a future with middle-of-the-road mitigation and adaptation to climate change (Fricko et al. 2017).

An initial set of potentially useful environmental parameters representing current conditions were obtained from the NZESM for the midpoint of every 1x1 km cell within the New Zealand region (see Section 2.5.3). The NZESM uses two different ocean model grids. Physical variables are available on a high-resolution 0.2° grid (~ 17km around New Zealand) and on a 1° (~90 km around New Zealand), while ocean biogeochemistry is available on the 1° degree only. To be consistent between ocean physics and ocean biogeochemistry only data from the 1° grid was used for this project. Both horizontal grids have 75 vertical levels with a thickness of 1m at surface and increasing with depth to about 250m. The bottom grid cell is partially filled to allow for an improved representation of bottom topography and properties near the sea floor. The interpolation of all 2-D fields (such as seafloor temperature and benthic nitrogen) for each grid point of the 1 km target grid was done through bilinear interpolation of all surrounding model grid cells, after extrapolating ocean values onto land. This allows for a better representation of the near-shore values. For the interpolation of 3-D seafloor variables, a similar approach was applied, but with consideration of vertical model grid and model bathymetry. Here values at the nearest vertical level of all surrounding model grid cells were used for the bilinear interpolation, except when they included land; in these cases, bottom values of the grid cell were used.

In addition to the NZESM variables, some potentially important environmental parameters that vary only spatially (rather than temporally) were also compiled. These comprised: seabed depth, slope, seamount, and aspects of seabed roughness and substrate composition. Spatial grids of these parameters were obtained from published research and databases, and NIWA records. Updated data layers of seafloor substrate composition for the New Zealand region have recently been produced (Bostock et al 2018a, 2018b). The availability of these new sediment data layers (representing concentrations of sand, mud, and carbonate) for use in the habitat suitability models is important because substrate is likely to have a substantial influence on the ability of corals to settle and grow. Nonetheless, it is worth noting here that these data layers are gridded at 1 km, a spatial scale that is typically much larger than the scale at which substrate type influences sessile fauna such as corals, and their formation required a high degree of interpolation due to the uneven distribution of sediment sampling in the region.

The full initial set of environmental predictors is shown in Table 2-2. This set was subsequently reduced to a core set suitable for habitat suitability models in which too many variables can lead to overfitting and highly correlated variables can adversely affect model performance and interpretation (e.g., Huang et al. 2011). The initial set of 19 variables was reduced in a stepwise manner by excluding the variable most highly correlated with the remainder. This was done by calculating the variance inflation factor (VIF) for each variable using the R library HH (Heiberger, 2011), following the method described by Yesson et al. (2015). Firstly, values for each variable were determined at the unique locations of all coral presence records combined, and VIFs calculated. The variable with the highest VIF was then discarded and the process repeated with the remaining variables, continuing iteratively until all variables showed a VIF of less than 5, reflecting a low level of co-linearity (Heiberger and Holland, 2004).

Following the initial reduction of predictor variables, trial BRT models were run for each taxon using the reduced set of variables to assess their predictive power and further eliminate any unsuitable predictors. This was done using the BRT *gbm.simplify* function (Elith et al. 2006), which carries out a backwards stepwise elimination of variables. This simplification process firstly assesses the relative contributions of each variable in terms of deviance explained, with the lowest contributing variables removed from the model, before the model is refitted with the remaining environmental variables. The change in deviance explained that resulted from removing the variable was then examined and the process repeated until all variables were sequentially removed. The final models were created by refitting the model with a reduced variable set that balanced the deviance explained with the number of predictor variables (after Stephenson et al. 2020) which were then used to construct final models for both BRT and RF methods. This procedure aims to provide a balance between providing models with as much explanatory power as possible and minimising fitting of models to noise rather than signal.

Table 2-2: Initial set of environmental predictors considered for HSM models.

Variable	Description	Units	Reference
ESM variables			
Nitrogen	Benthic Nitrogen (BEN_N) [2-D]	mmol/m^2	NZESM
Detrital flux	Total detrital flux to seabed (DETFLUX3) [3-D]	mmol/m^3	NZESM
DIC	Dissolved Inorganic Carbon (DIC_C) [3-D]	mmol/m^3	NZESM
POC	Particulate Organic Carbon flux (EXPC3) [3-D]	mmol/m^3	NZESM
Calcite	Calcite saturation state (OM_CAL3) [3-D]	–	NZESM
Oxygen	Dissolved oxygen at the seafloor (OXY) [3-D]	mmol/m^3	NZESM
Pressure	Seawater pressure at seafloor (PBO) [2-D]	dbar	NZESM
Aragonite	Aragonite saturation state (SFR_OARG) [2-D]	–	NZESM
Salinity	Seafloor salinity (SO) [2-D]	psu	NZESM
Temperature	Seafloor temperature (TO) [2-D]	Degrees C	NZESM
Fixed variables			
Seamount	Seamount positions	–	Rowden et al. (2008), Mackay (2007)
Carbonate	Percentage of carbonate in sediment	%	Bostock et al. 2018ab
Mud	Percentage of mud in sediment	%	Bostock et al. 2018ab
Sand	Percentage of sand in sediment	%	Bostock et al. 2018ab
BPI-broad	Benthic Position Index – broad-scale	–	NIWA
BPI-fine	Benthic Position Index – fine-scale	–	NIWA
Depth	Seabed depth	m	NIWA
Slope-percent	Seabed slope in percent	%	NIWA
Stdev-slope	Standard deviation of seabed slope	–	NIWA

2.4 Model types

Random Forest models

Random Forest (RF) models are a non-parametric approach which build classification or regression trees using random subsets of the input data (Breiman 2001). The RF models were built in R (R Core Team 2018) using *randomForest* and tuned using the train function in *caret* to select optimal values for complexity parameters *mtry* (the number of variables used in each tree node), *maxnodes* (the maximum number of terminal nodes in each trees), and *ntree* (the number of trees to grow). The RF approach to habitat suitability modelling has also been successfully applied in the past to benthic invertebrate data in the New Zealand region (Georgian et al. 2019, Rowden, et al. 2017).

Boosted Regression Tree models

Boosted Regression Tree (BRT) models are an advanced form of additive regression based on decision trees, where the individual terms of the regression are simple trees, fitted in a stage-wise manner. Short trees are formed by relating a response to recursive binary splits of the data, then combined (boosted) to improve predictive power by focussing each successive tree on model residuals. Tree-based methods such as BRT and RF have the advantage over traditional methods that they can easily handle missing data, outliers, categorical as well as continuous variables, and automatically handle interactions between predictors (Elith et al. 2008). Detailed descriptions of the BRT method are available in Ridgeway (2019) and Elith et al. (2008). The BRT method has been widely used in ecological applications and has performed well in previous studies of deep-water invertebrate and fish distributions in New Zealand (e.g., Compton et al. 2013, Georgian et al. 2019, Leathwick, et al. 2006, Rowden et al. 2017, Tracey et al. 2011).

All BRT analyses were undertaken in R using the *dismo* package (Hijmans et al. 2011). Tree-complexity (number of splits) was set to 3, allowing for a level of variable interaction, and the learning rate (which determines the weight given to each successive tree in the model) was set to 0.001, but adjusted where necessary so that the number of trees in the final models exceeded 1000; a bag fraction of 0.6 was used with a random 10-fold cross evaluation method following guidelines in Elith et al. (2008).

Ensemble models

We produced an ensemble model for each taxon, to incorporate the predictions and underlying assumptions and modelling strategies of each model type. This approach limits dependence on a single model type or structural assumption and enables a more robust characterization of the predicted spatial variation and uncertainties (Robert et al. 2016). Ensemble models were constructed by taking weighted averages of the predictions from the BRT and RF models, using methods adapted from Oppel et al. (2012), Anderson et al. (2016), Rowden et al. (2017), and Georgian et al. (2019). This adapted procedure derives a two-part weighting for each component of the ensemble model, taking equal contributions from the overall model performance (cross-validated AUC values) and the uncertainty measure (CV) in each cell (see Section 2.5), as follows,

$$W_{BRT1} = \frac{MPS_{BRT}}{MPS_{BRT}+MPS_{RF}} \text{ and } W_{RF1} = \frac{MPS_{RF}}{MPS_{BRT}+MPS_{RF}}$$

$$W_{BRT2} = 1 - \frac{CV_{BRT}}{CV_{BRT}+CV_{RF}} \text{ and } W_{RF2} = 1 - \frac{CV_{RF}}{CV_{BRT}+CV_{RF}}$$

$$W_{BRT} = \frac{W_{BRT1}+W_{BRT2}}{2} \text{ and } W_{RF} = \frac{W_{RF1}+W_{RF2}}{2}$$

$$X_{ENS} = X_{BRT} * W_{BRT} + X_{RF} * W_{RF}$$

$$CV_{ENS} = \sqrt{\frac{(CV_{BRT} * X_{BRT})^2 * W_{BRT}^2 + (CV_{RF} * X_{RF})^2 * W_{RF}^2}{X_{ENS}^2}}$$

where MPS_{BRT} and MPS_{RF} are the model performance statistics; X_{BRT} and X_{RF} are the model predictions; CV_{BRT} and CV_{RF} are the bootstrap CVs; and X_{ENS} and CV_{ENS} are the weighted ensemble predictions and weighted CVs, respectively, from which maps of predicted species distribution and model uncertainty were produced.

2.5 Model performance and outputs

2.5.1 Model performance

Model performance was evaluated for the BRT and RF models primarily using the AUC metric (area under the Receiver Operating Characteristic curve), which for these types of models are defined as the area under a plot of the fraction of true positives versus the fraction of true negatives. In general, AUC scores over 0.5 indicate better than random performance, values over 0.7 indicate adequate performance, and values over 0.8 indicate excellent performance (Hosmer et al. 2013). Several additional performance metrics were also calculated; sensitivity (% of presences correctly identified), specificity (% of absences correctly identified); True Skill Statistic (TSS, which takes into account specificity and sensitivity to provide an index ranging from -1 to +1, where +1 equals perfect agreement and -1 = no better than random, Allouche et al. (2006)). In addition, a cut-off value based on maximising sensitivity+specifity was calculated, in order to predict a binary outcome from the model (i.e. presence or absence).

These performance metrics were assessed for each taxon and model by a process of cross-validation, using a procedure which also considers spatial autocorrelation in the input data (Roberts et al. 2017, Valavi et al. 2019). For this, model input data for each taxa was partitioned systematically into 10 spatial blocks - the size of which were based on a distance for which the residuals of a BRT model were approximately independent, as assessed by a variogram – to create training and test data sets, with 9 blocks used to construct a preliminary model which was then applied to the remaining block to measure performance. This process was repeated ten times, leaving out each of the 10 blocks in turn, for each model type, and average AUC, sensitivity, specificity, and TTS scores calculated to represent overall performance.

2.5.2 Model precision

To assess the relative confidence in predictions across the model extent, we used a bootstrap technique to produce spatially explicit uncertainty measures, after Anderson et al. (2016) and Georgian et al. (2019). Random samples of the presence data were drawn, with replacement, from the model input datasets and to these were added an equal number of absence records, drawn randomly from the master coral occurrence dataset. Presence-absence models of each type were then constructed using the same settings as the originals, and predictions of habitat suitability made for each cell of the model extent. This process was repeated 200 times for each model type (BRT and RF) resulting in 200 estimates of abundance for each taxon in each cell. Model uncertainty in each cell was then calculated as the coefficient of variation (CV) of the bootstrap output.

2.5.3 Model outputs

Model predictions were limited to the standard New Zealand Region (CANZ 2008), extending from 157° E to 167° W and 24° to 57.5° S, an area that encompasses the New Zealand EEZ and TS. There was no upper limit to the depth range modelled, although there were few presence points for many taxa near to the shorelines and lack of environmental data points near the shore required a high degree of interpolation for some predictor variables. The maximum depth for the models was set to 2000 m, as few taxonomic records were available below this level. The resolution (cell size) of the environmental input grids and model output maps was set to 1 km^2. All resulting maps are presented using the Albers Equal Area projection centred at 175°E and 40°S (EPSG::9191), a standard format now accepted by DOC and Fisheries New Zealand (FNZ) (Wood et al. in prep). Maps are overlaid on a hill-shade layer to illustrate variations in bathymetry.

Because absence data used in the model were obtained randomly from a large set of sampling stations and do not represent true absence, results are presented as relative habitat suitability and not as probability of presence. Maps of model precision take into account variability in the model input data, and not any other sources of error.

2.6 Overlap with trawl fishing effort

A spatial grid of the combined aggregated swept area by bottom trawling for inshore and deepwater target species, covering recent fishing years only (2007–08 to 2017–18), was compiled from annual layers calculated by Baird and Mules (2019) and provided by FNZ. These data are limited to the New Zealand Territorial Sea (TS) and EEZ.

To compare the distribution of bottom trawl fishing effort with habitat suitability and identify the areas at greatest risk from contact with fishing gear, cells in the top 25% of aggregated swept area and also in the top 25% of habitat suitability (excluding regions outside of the TS and EEZ) were identified for each taxon. These regions were plotted and the number of overlapping 1x1 km cells enumerated.

3 Results

3.1 Selected taxa

Habitat suitability models were produced for all 12 taxonomic groups listed in Table 2-1. This was made possible due to improved efficiencies in coding and the automation of various parts of the model development. As a result of this, models were able to be produced for the first time for the alcyonacean genera *Primnoa* and *Corallium*, and the anthoathecatan genera *Errina* and *Stylaster*.

3.2 Predictor variable selection

The variance inflation factor procedure reduced the initial 19 potential predictor variables to a smaller set with minimal correlations among them. These comprised *Carbonate*, *Mud*, *Sand*, *BPI-fine*, *Slope-percent*, *Seamount*, *Nitrogen*, *Detrital flux*, *Oxygen*, *Depth*, and *Salinity*. Despite a level of correlation with some of these variables, two additional variables were included with importance for skeletal composition in most corals; *Calcite* and *Aragonite*. Aragonite is the polymorph of calcium carbonate used in the skeletons of scleractinian and anthoathecatan corals whereas calcite is used in antipatharian and alcyonacean corals (Bostock et al. 2015). As such, only one of either *Calcite* or *Aragonite* were included in subsequent model development, depending on the taxon being modelled. Refer to Figure A-1 toFigure A-6 for maps illustrating the current and future spatial variability of each predictor used in the models.

Backwards stepwise BRT models based on these individualised sets of variables further reduced the number of variables from use in final model construction to between 4 and 8 (see Table 3-2).

3.3 Model performance

Tuning of Random Forest models to select optimal settings did not result in clearly improved fits, based on the “out-of-bag” estimate of error rate, therefore default settings for *mtry* (square root of the number of variables), *ntree* (501), and *maxnodes* (1) were used in the models for each taxon.

Cross-validated AUC scores for the individual BRT and RF models ranged from 0.76 (*Goniocorella dumosa* - BRT) to 0.93 (*Corallium* spp. – RF), indicating adequate models (AUC > 0.7) in all cases (Table 3-1). AUC scores for most taxa, for both methods, were between 0.8 and 0.9. Only for *Goniocorella dumosa* was AUC less than 0.8 for both model types. The best fit by this measure was for *Corallium* spp. where AUC was 0.92–0.93.

Sensitivity and specificity values were also well above 0.5 in all cases. Sensitivities were over 0.9 for *Enallopsammia rostrata, Corallium* spp., and both black coral genera. Specificities were mostly slightly lower than Sensitivities, reaching a maximum of 0.86–0.88 for *Corallium* spp. and *Errina* spp.

True Skill Statistic (TSS) values were all at acceptable levels, all much greater than zero. The highest values were again for *Corallium* spp. with the black coral models and *Errina* spp. model also performing well by this measure. Cut-off values for presence-absence ranged from a low of 0.27 (*Leiopathes* spp. – RF) to a high of 0.52 (*Paragorgia* spp. (BRT) and *Corallium* spp. (RF)) and were in most cases slightly lower in the RF models than in the BRT models.

For most taxa most of the performance metrics were similar between the two model types, suggesting similar fits to the data. This was more noticeable for the taxa with better fitting models, e.g., *Corallium* spp. and *Errina* spp., than for taxa with weaker performance metrics, e.g., *Stylaster* spp. and *Primnoa* spp.

Table 3-1: Performance metrics for the BRT and RF components of coral Habitat Suitability Models.

Taxon	AUC		Sensitivity		Specificity		TSS		Cutoff	
	BRT	RF	BRT	RF	BRT	RF	BRT	RF	BRT	RF
Enallopsammia rostrata	0.87	0.83	0.92	0.84	0.79	0.81	0.71	0.65	0.42	0.40
Solenosmilia variabilis	0.84	0.82	0.88	0.78	0.78	0.82	0.66	0.60	0.43	0.40
Goniocorella dumosa	0.76	0.77	0.85	0.87	0.64	0.67	0.48	0.55	0.41	0.28
Madrepora oculata	0.79	0.85	0.83	0.86	0.80	0.83	0.63	0.69	0.43	0.36
Paragorgia spp.	0.80	0.77	0.80	0.82	0.82	0.73	0.62	0.55	0.52	0.38
Primnoa spp.	0.77	0.83	0.85	0.91	0.80	0.85	0.65	0.75	0.40	0.33
Corallium spp.	0.93	0.93	0.95	0.95	0.88	0.87	0.82	0.82	0.49	0.52
Keratoisis spp. & *Lepidisis* spp.	0.82	0.82	0.81	0.85	0.88	0.71	0.70	0.56	0.58	0.29
Bathypathes spp.	0.86	0.83	0.89	0.87	0.83	0.80	0.72	0.67	0.46	0.39
Leiopathes spp.	0.86	0.85	0.84	0.94	0.86	0.77	0.70	0.71	0.50	0.36
Errina spp.	0.89	0.90	0.87	0.85	0.87	0.88	0.74	0.74	0.50	0.50
Stylaster spp.	0.79	0.84	0.72	0.88	0.65	0.77	0.57	0.66	0.37	0.40

Dissolved *Oxygen* was the most influential variable overall across the 12 taxa modelled, ranking first in importance in both the RF and BRT models for the bamboo corals and for both black coral models; *Oxygen* also ranked first in the RF model for *Madrepora oculata*. The models for *Stylaster* spp. were the exception, with *Oxygen* eliminated for this taxon during the variable selection process (Table 3-2). *Nitrogen* was also strongly influential in many of the models – ranked most important for *Primnoa* spp. and *Corallium* spp. models of both types and ranked second or third in many other models – and was eliminated only from the bamboo coral models. Seafloor *Depth* was present in only 7 of the 12 models but was ranked most important in both model types for *Goniocorella dumosa* and second most important for *Errina* spp. For the four branching scleractinian corals the key variable overall was *Aragonite,* ranking first or second in each model. For the calcite dependent taxa *Calcite* was present in all models except for *Corallium* spp. and was ranked first for *Paragorgia* spp. and second for bamboo coral models. Of the remaining variables, *Mud* was top ranked for *Errina* spp., and *BPI-fine* was top ranked for *Stylaster* spp.; the least important variables overall, with only low ranking in a few of the models, were *Salinity*, *Seamount*, and *Slope-percent*.

Table 3-2: Variable importance ranks by taxon and model type (RF/BRT). The rank signifies the order of variable importance in each model; variables are ordered from left to right by mean overall importance across the 12 taxa.

	Oxygen	*Nitrogen*	*Depth*	*Aragonite*	*Mud*	*Calcite*	*BPI-fine*	*Carbonate*	*Detrital flux*	*Salinity*	*Seamount*	*Slope-percent*
Enallopsammia rostrata	2/2	3/4	4/5	1/1	5/3					6/6		
Solenosmilia variabilis	5/5	2/2	3/3	1/1				4/4				
Goniocorella dumosa	6/3	5/6	1/1	2/2	4/8			7/4	3/5			8/7
Madrepora oculata	1/3	3/5	5/4	2/1				4/2				
Paragorgia spp.	3/2	2/3				1/1	5/5		4/4			
Primnoa spp.	3/2	1/1			4/4	2/3		6/5	5/6			
Corallium spp.	3/3	1/1	4/5				2/2		5/4		6/6	
Keratoisis spp. & *Lepidisis* spp.	1/1		4/4		3/3	2/2						
Bathypathes spp.	1/1	2/2	5/6		6/5	3/4	4/3					
Leiopathes spp.	1/1	5/5	3/2			2/4	4/3			6/6		
Errina spp.	3/4	4/3	2/2		1/1			5/6			6/5	
Stylaster spp.		5/4	4/2	3/7	2/3		1/1	7/6		6/5		

3.4 Predicted coral habitat distributions

Maps of predicted habitat suitability produced by BRT and RF models showed similar overall patterns but with some subtle differences apparent in some regions for most taxa, as illustrated below for two representative genera *Paragorgia* spp. and *Leiopathes* spp. (Figure 3-1). For *Paragorgia* spp. RF-based maps show a slightly higher level of suitability than BRT-based maps in most regions, most noticeable south of the Chatham Rise. The overall level of habitat suitability is similar between models for *Leiopathes* spp. but is noticeably greater in the north for the BRT-based map. With generally similar AUC scores for most species, maps for the weighted ensemble models essentially represent an average of the BRT and RF model outputs.

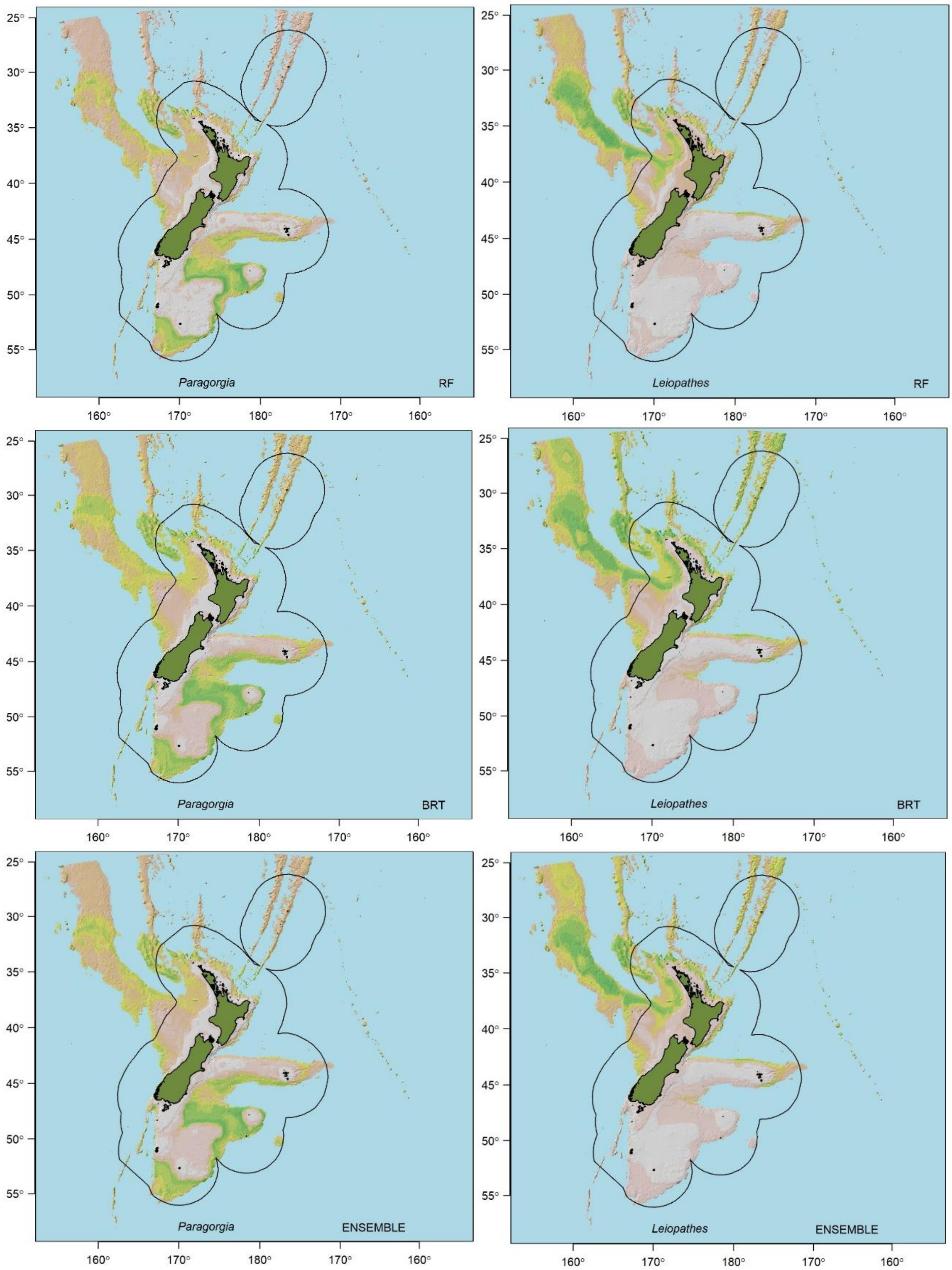


Figure 3-1: Comparison of RF, BRT, and ensemble models for two representative taxa. Left, *Paragorgia* spp.; right, *Leiopathes* spp.; top, RF; middle, BRT; bottom, ensemble models.

3.4.1 Scleractinia

Patterns of habitat suitability showed considerable variation among the four reef-forming scleractinian corals, both spatially and temporally (Figure 3-2 to Figure 3-5). Historical distributions of *Enallopsammia rostrata* have been focussed on deeper regions, especially on the rises and ridges north of New Zealand and around the edges of the Chatham Rise and sub-Antarctic plateaus. Predicted future distributions are very similar in the north but indicate potentially improved habitat suitability into shallower regions of the Challenger Plateau, Chatham Rise, and sub-Antarctic plateaus. Precision is also generally greater in southern regions and tends to be lowest in shallower waters around the New Zealand coastline.

The predicted distribution of *Solenosmilia variabilis* overlaps strongly with that of *Enallopsammia rostrata* with a focus on deeper regions in the north and around the southern rises (Figure 3-3). Future distributions are predicted to change little on the northern ridges and rises, but on the Challenger Plateau, Chatham Rise, and sub-Antarctic, are predicted to slightly increase in shallower regions and decrease in deeper regions. Precision varies considerably for this species, both spatially and temporally, but is notably low nearer the coast and in the future in most areas.

The predicted distribution of *Goniocorella dumosa* is much shallower than that of *Enallopsammia rostrata* and *Solenosmilia variabilis*, with greatest habitat suitability on shallower parts of the Chatham Rise, sub-Antarctic, and continental shelf around the North and South Islands (Figure 3-4). In the north, high levels of habitat suitability are limited to small regions of shallower seafloor. There is predicted to be only subtle changes in habitat suitability in most regions by the end of the century. Estimated precision of these distributions is relatively high, especially in shallower parts, with CVs of less than 40% over much of the model extent and little difference between historic and future patterns.

The distribution of *Madrepora oculata* is similar to that of *Enallopsammia rostrata* and *Solenosmilia variabilis*, with the highest levels of habitat suitability strongly associated with the underseas features in the north and around the margins of those in the south (Figure 3-5). As with each of the other branching coral species, predicted changes in future habitat suitability are more pronounced in southern regions than northern, with shallower water becoming slightly more suitable across the main rises and plateaus around New Zealand and deeper water becoming slightly less suitable in a few areas, such as eastern Campbell Plateau. Precision around these predictions is mostly less than 40% and relatively constant across the modelled extent, and notably lower across the Challenger Plateau in the future compared to historic.

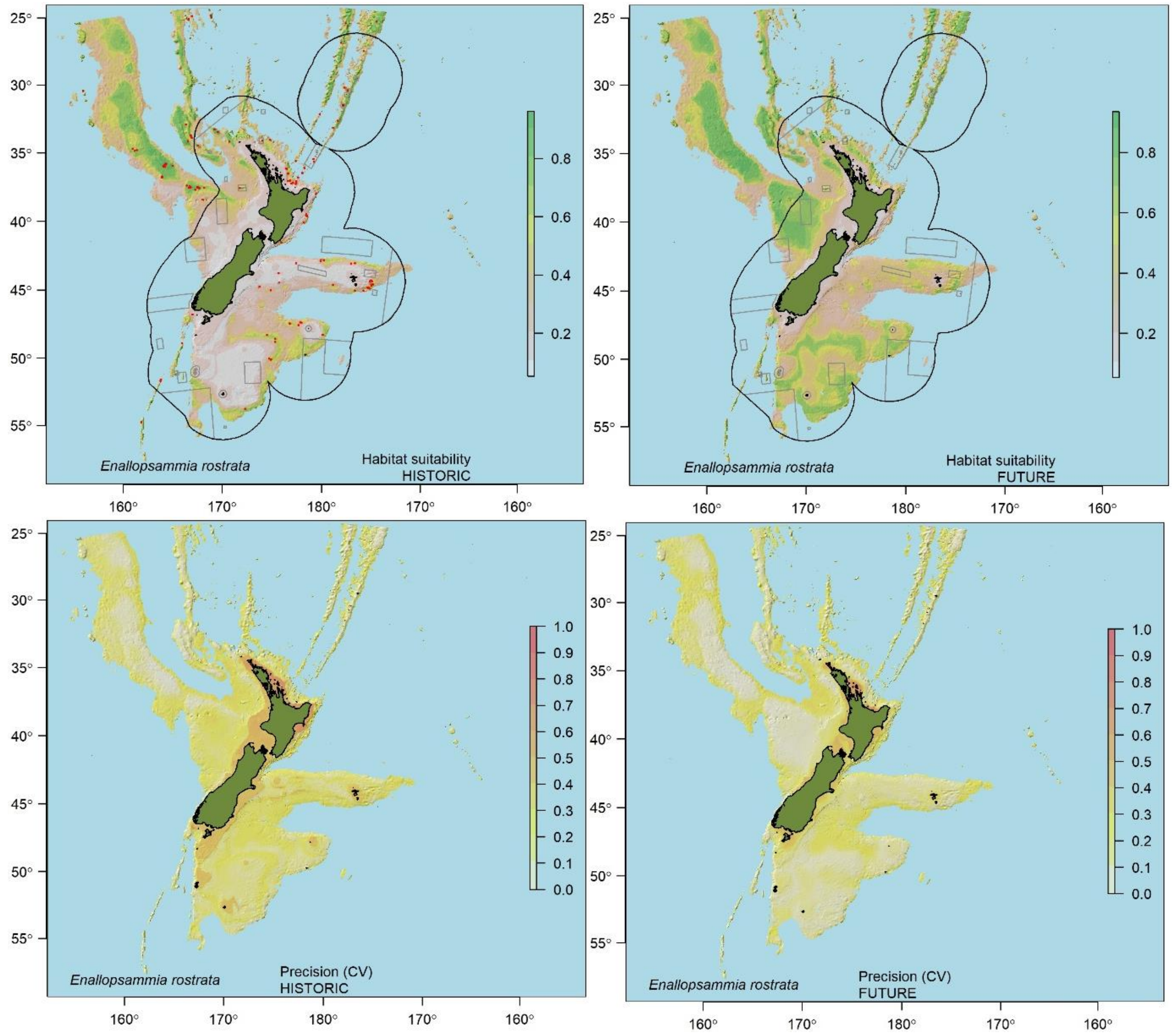


Figure 3-2: ***Enallopsammia rostrata*. Predicted habitat suitability.** Left, historic distribution (red dots = presence locations); Right, future distribution; Top, relative habitat suitability; Bottom, precision (CV). The boundaries of the New Zealand EEZ (black lines) as well as Benthic Protection Areas, Large Marine Reserves, and seamounts closed to bottom trawling (grey lines) are provided in the upper panels.

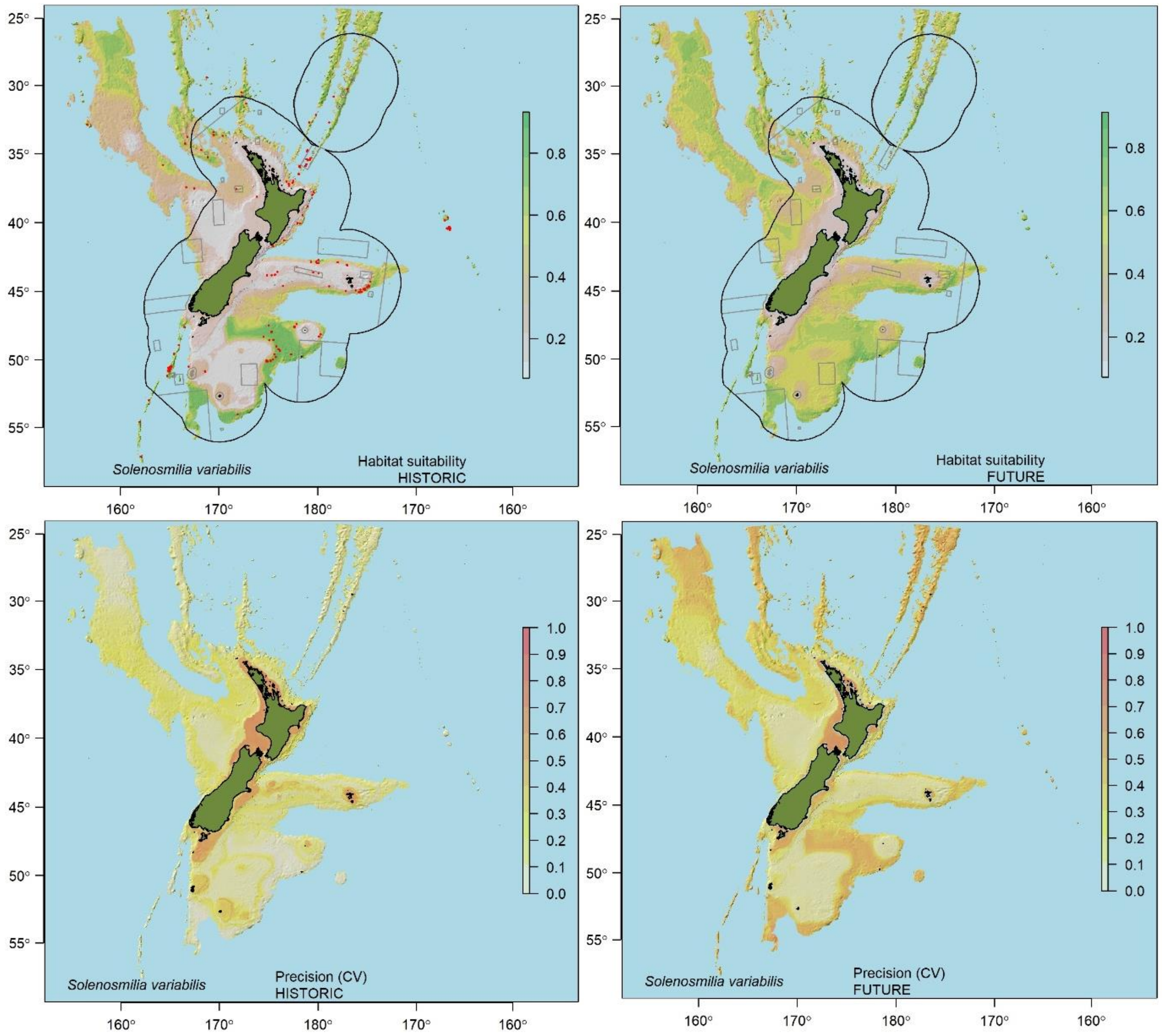


Figure 3-3: ***Solenosmilia variabilis*. Predicted habitat suitability.** See Figure 3-2 caption for further details.

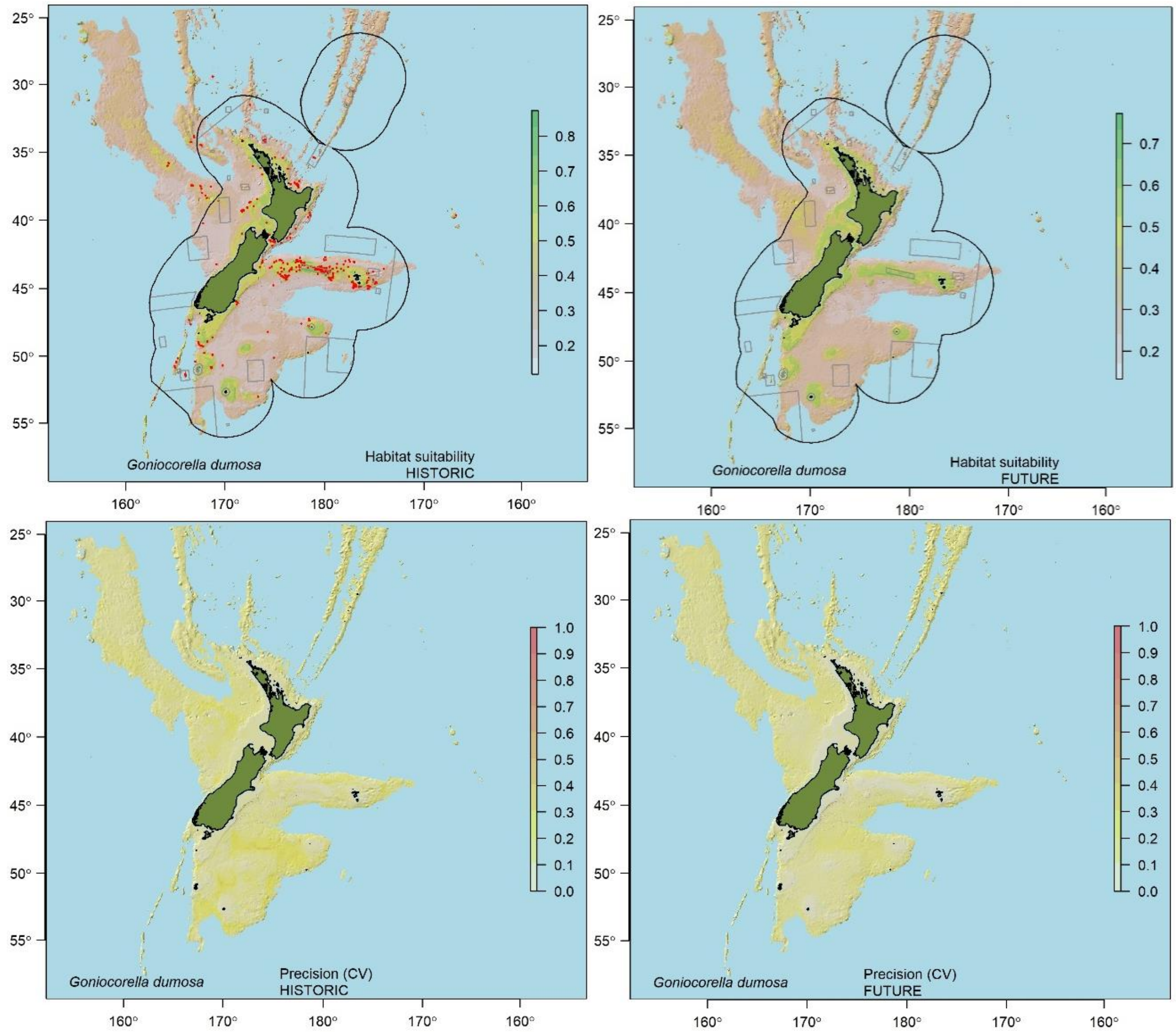


Figure 3-4: ***Goniocorella dumosa*. Predicted habitat suitability.** See Figure 3-2 caption for further details.

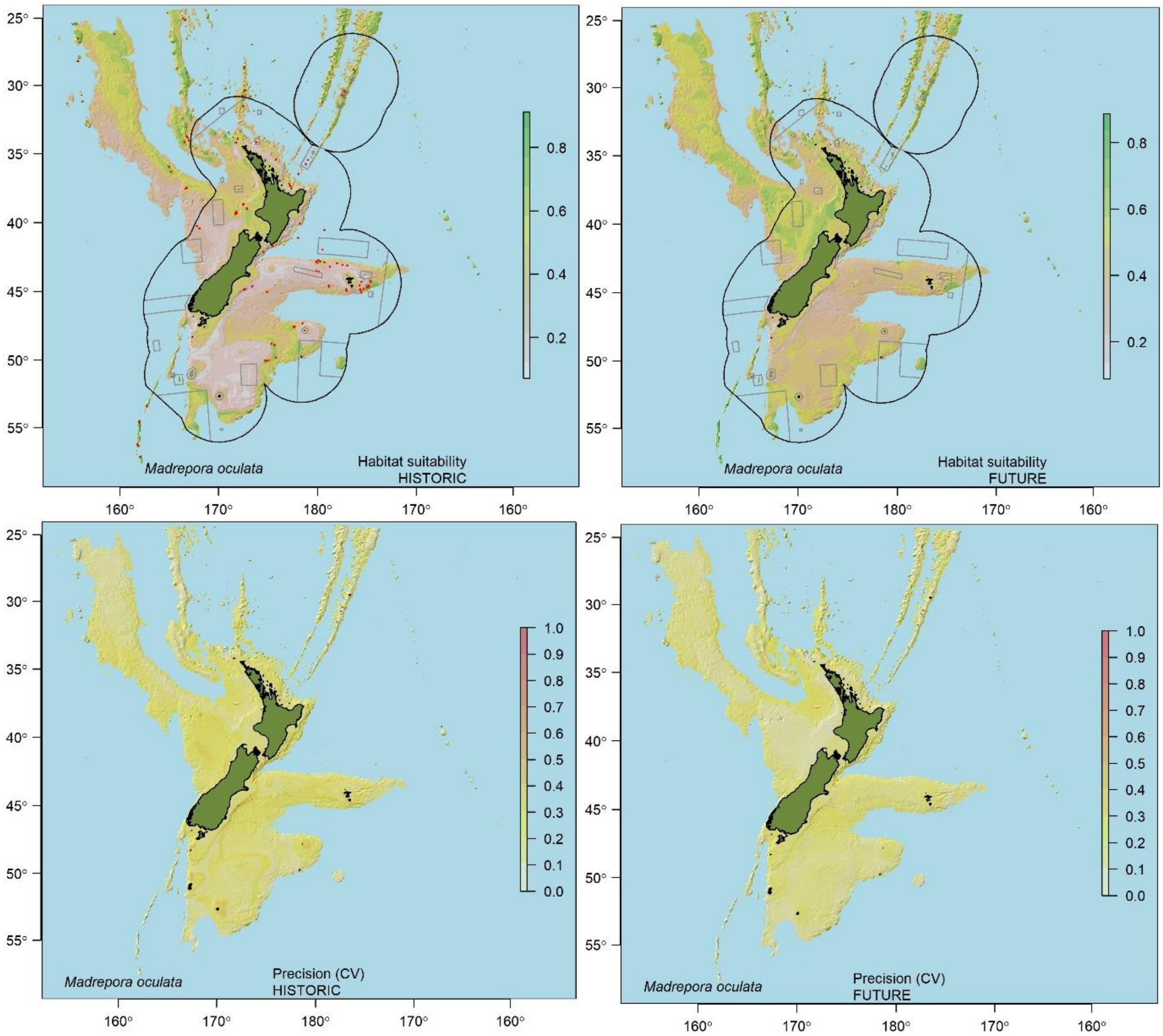


Figure 3-5: ***Madrepora oculata*. Predicted habitat suitability.** See Figure 3-2 caption for further details.

3.4.2 Alcyonacea

Habitat suitability of *Paragorgia* spp. is uniformly quite low north of about 30° S, and this is not predicted to change in the future (Figure 3-6). Elsewhere, the highest suitability for this genus has been around the West Norfolk Ridge and deeper fringes of the southern Chatham Rise and Campbell Plateau. Habitat suitability is predicted to generally decrease in these regions over time, with a slight shift into shallower depths indicated for the Campbell Plateau. Precision is relatively constant across the modelled extent although notably lower around the coastlines of the main islands.

Predicted habitat suitability of *Primnoa* spp. is similar to that of *Paragorgia* spp., with high suitability around the deeper areas of the Campbell and Bounty Plateaus and low over much of the Challenger Plateau and northern Chatham Rise (Figure 3-7). There is expected to be a slight shift of habitat suitability into shallower depths in the future, tending towards a more even distribution of suitability, and with lower suitability in some of the deeper regions. Again, there is little change in suitability expected for this species over time in regions north of about 30 °S. Precision tends to be greatest where suitability is greater, notably increasing across the shallower reaches of the Campbell Plateau over time.

For the *Corallium* spp. precious corals (Figure 3-8), habitat suitability is greatest in the north and is uniformly high across all depth ranges encountered north of about 30° S in the west and 35° S in the east. Suitability is low over much of the rises and plateaus around New Zealand, but there are areas of higher suitability in deeper water in the easternmost Chatham Rise, and deeper reaches of the Campbell and Bounty Plateaus. Habitat suitability is predicted to decrease in all the currently highly suitable areas and increase slightly in some shallower regions of the Challenger Plateau and Lord Howe Rise. Model precision tends to be greatest in areas of high habitat suitability and remains the same or increases for future predictions.

Patterns of habitat suitability for bamboo corals *Keratoisis* spp. and *Lepidisis* spp. are similar to that for *Corallium* spp. in many areas, but differ substantially on the southern Lord Howe Rise and outer Challenger Plateau where suitability is relatively high (Figure 3-9). Habitat suitability is predicted to generally decrease across the modelled region but improve slightly in shallower, currently poorly suitable, regions – especially on the Challenger Plateau. Model precision is relatively constant across the modelled region, more so and slightly lower overall for future predictions.

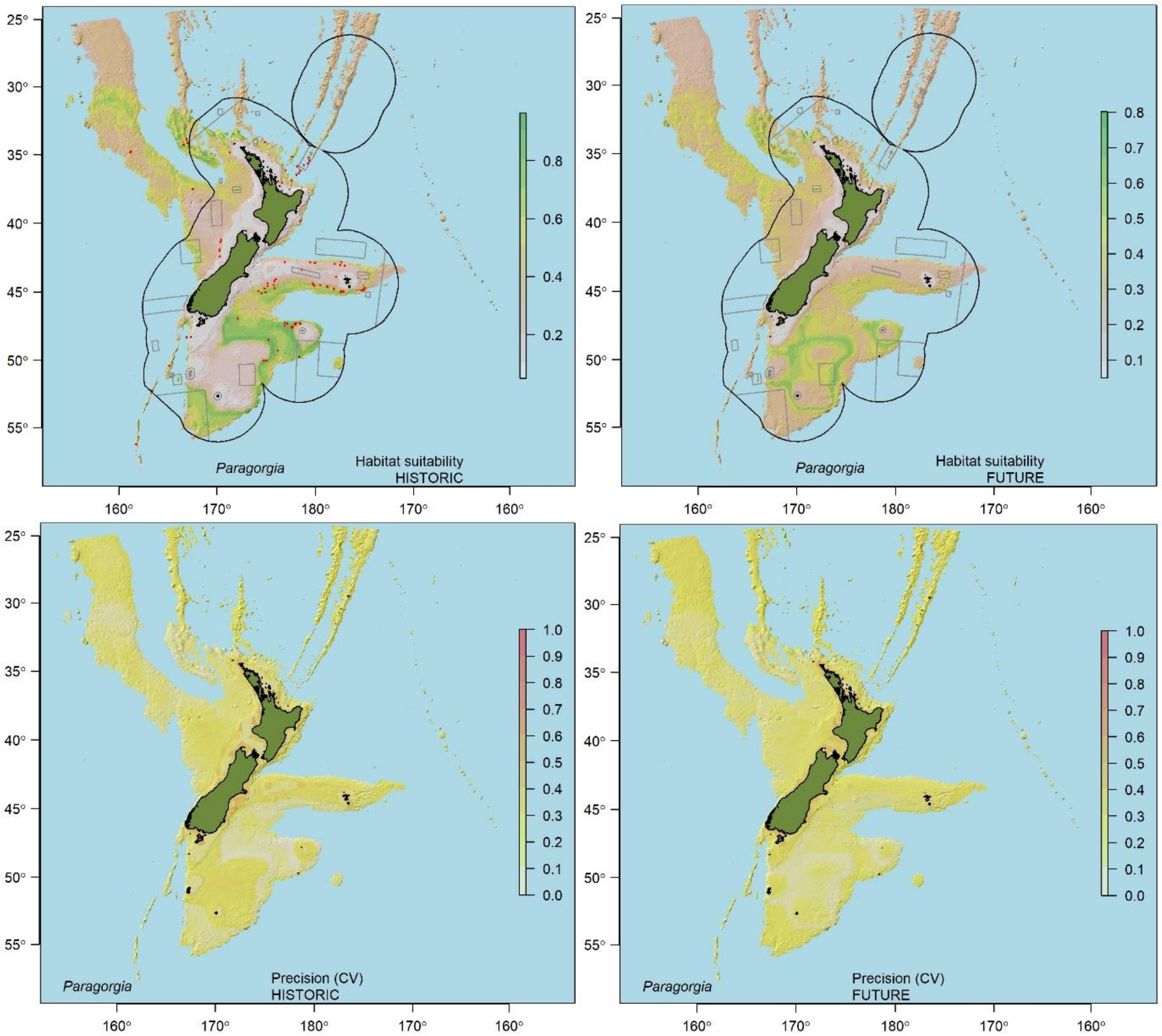


Figure 3-6: ***Paragorgia* spp. Predicted habitat suitability.** See Figure 3-2 caption for further details.

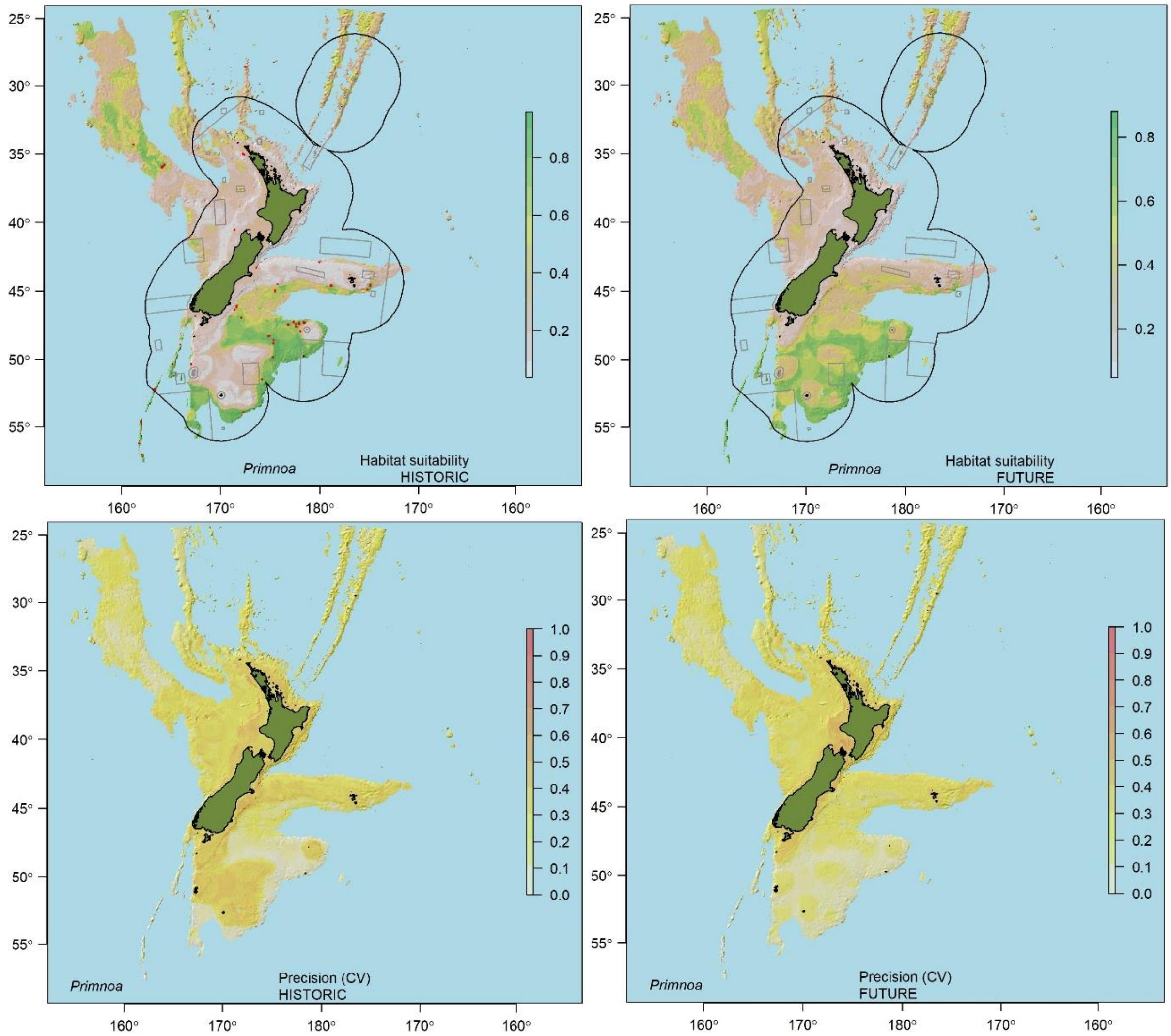


Figure 3-7: ***Primnoa* spp. Predicted habitat suitability.** See Figure 3-2 caption for further details.

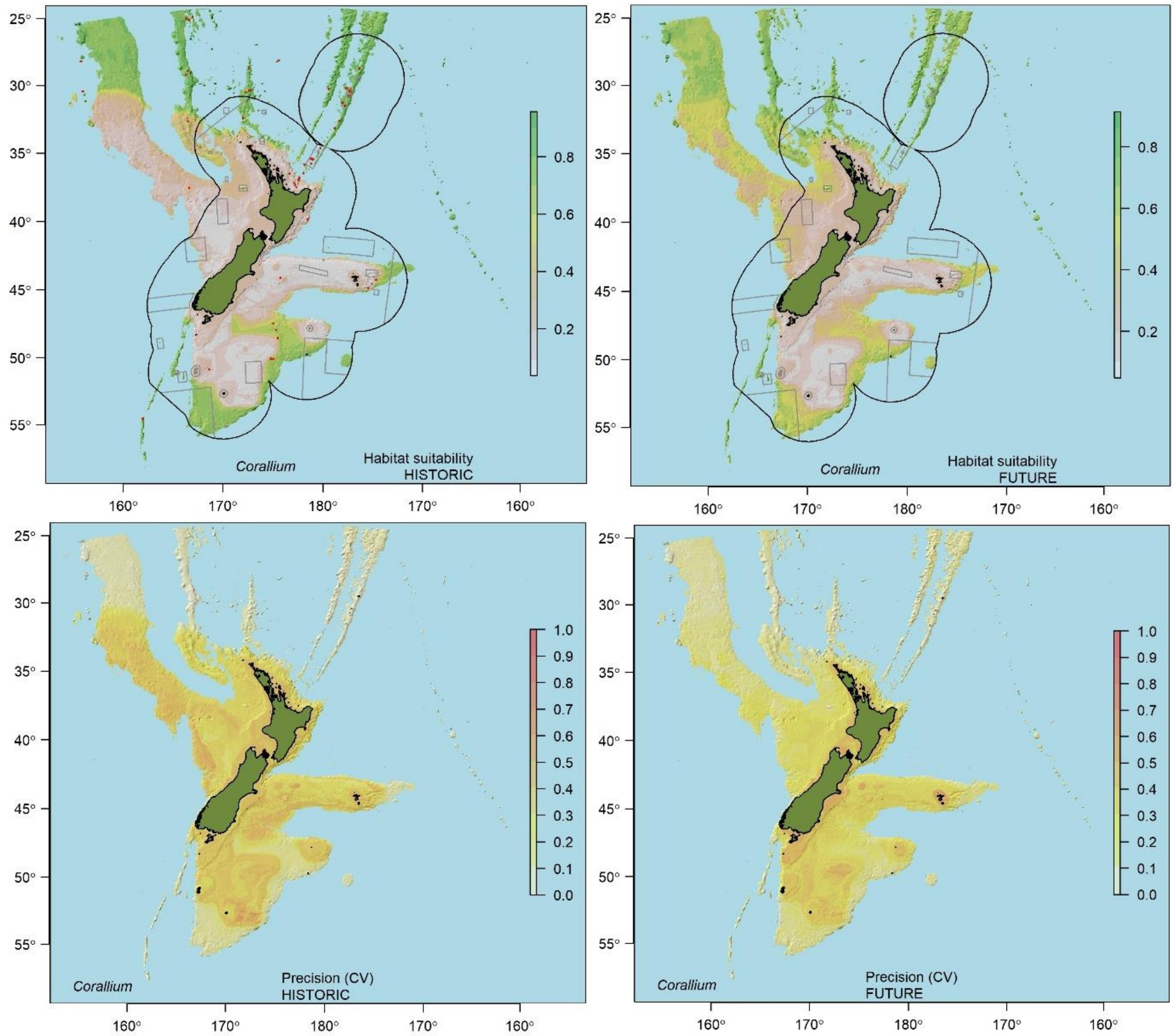


Figure 3-8: ***Corallium* spp. Predicted habitat suitability.** See Figure 3-2 caption for further details.

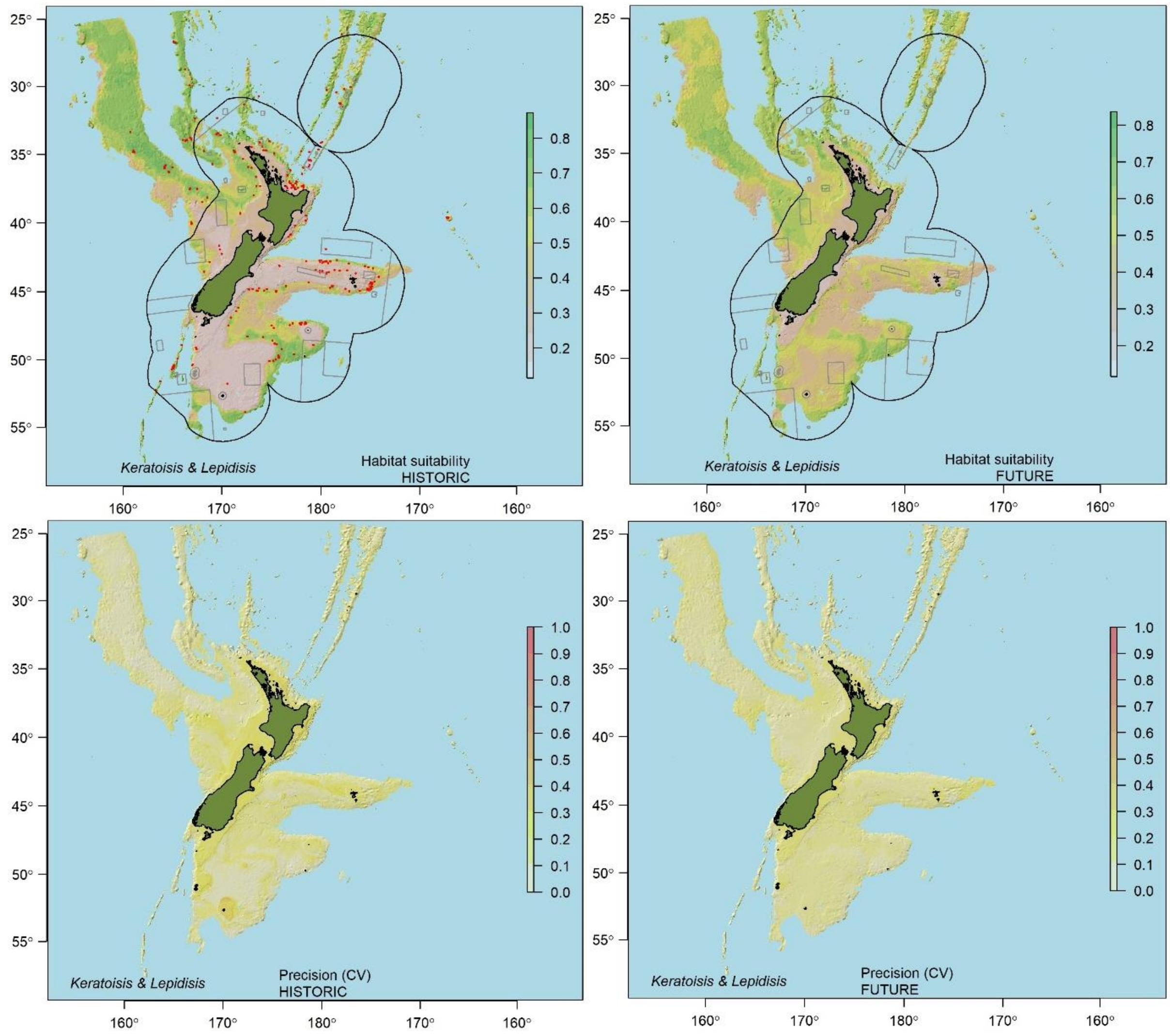


Figure 3-9: *Keratoisis* spp. and *Lepidisis* spp. Predicted habitat suitability.See Figure 3-2 caption for further details.

3.4.3 Antipatharia

The predicted current distribution of *Bathypathes* spp. is typical of most of the other coral groups, with highest suitability around the northern oceanic features, the deeper fringes of the rises and plateaus adjacent to mainland New Zealand, and the seamount chains and ridges in the east (Louisville) and south (Macquarie) (Figure 3-10). This distribution is predicted to remain relatively stable over time, with slightly decreasing suitability in the north and a subtle evening-out of suitability in the south. Precision is greatest in the northwest and lowest in coastal regions, and is more uniform for future predictions.

Habitat suitability predicts a more limited distribution for *Leiopathes* spp., with high levels of suitability centred around an area comprising the southern Lord Howe Rise, West Norfolk Ridge, and the margin of the continental shelf on northern Challenger Plateau and around the North Island (Figure 3-11). Models predict slightly lower suitability over much of this area in the future, but slightly increased suitability notable on the western Challenger Plateau. Small areas of high habitat suitability in deeper parts of the eastern Chatham Rise become slightly less suitable in the future.

Model precision, greatest in regions of highest and lowest suitability, is generally more uniform for future predictions.

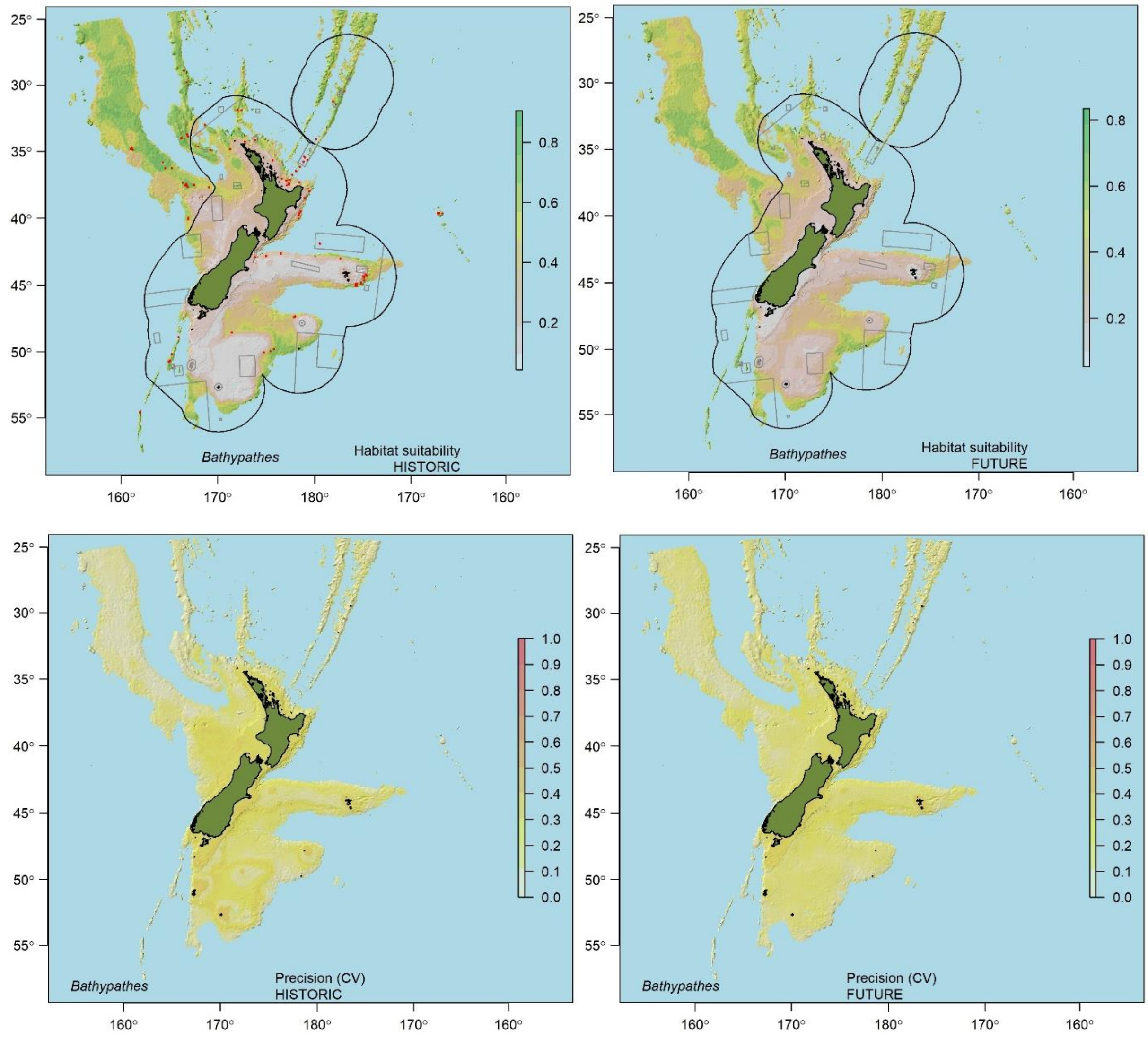


Figure 3-10: *Bathypathes* spp. Predicted habitat suitability. See Figure 3-2 caption for further details.

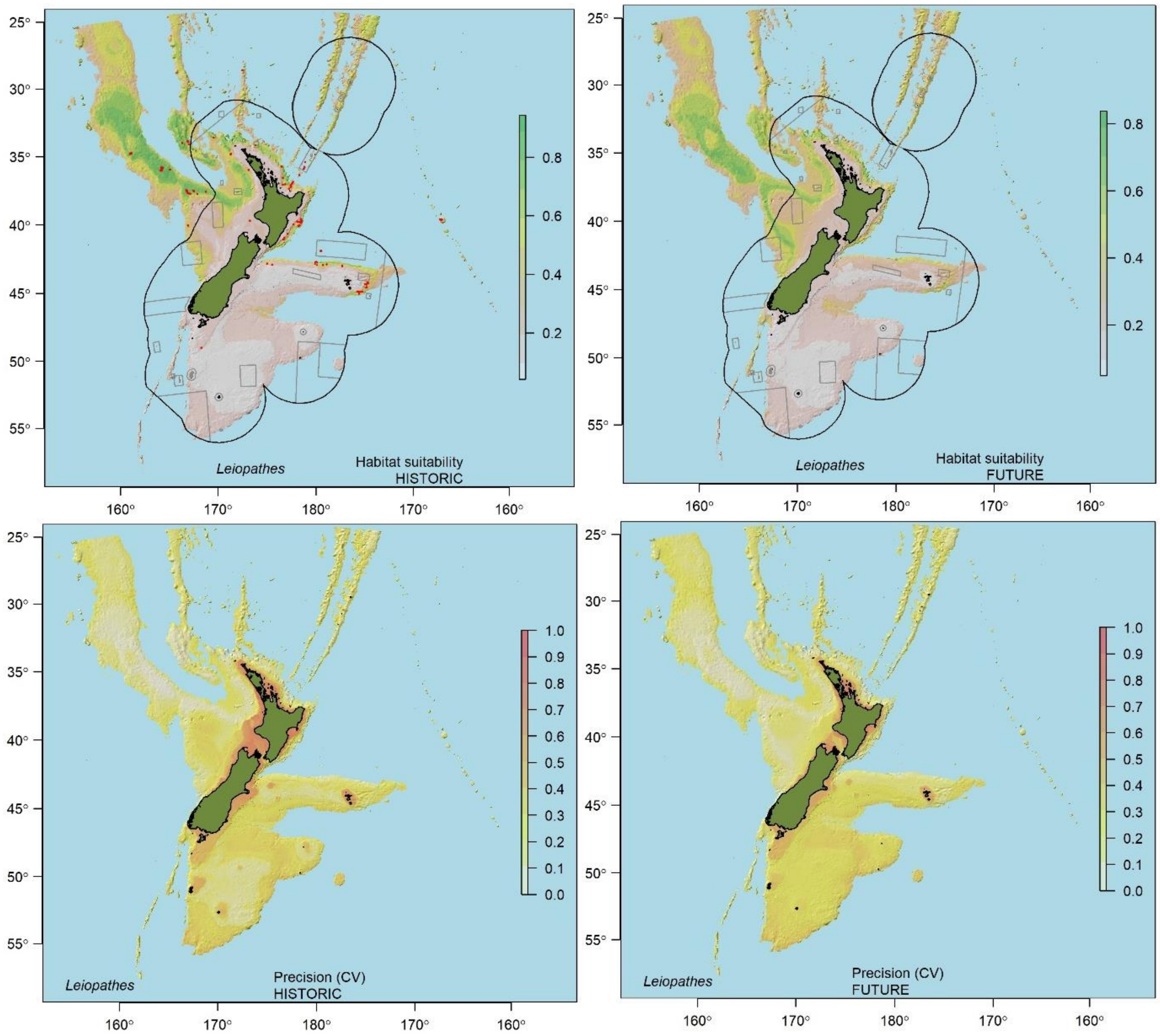


Figure 3-11: *Leiopathes* spp. Predicted habitat suitability. See Figure 3-2 caption for further details.

3.4.4 Anthoathecata

Models show a patchy distribution for the *Errina* spp. hydrocorals (Figure 3-12). High habitat suitability is indicated on the Stewart-Snares shelf, around all the sub-Antarctic islands, the Macquarie Ridge, eastern Campbell Plateau, and several small areas on the Chatham Rise and northern ridge features. This distribution is not predicted to alter substantially in any region in the future. Precision is high around many of the highly suitable areas, and moderate elsewhere, and differs little between the historic and future model predictions.

The *Stylaster* spp. hydrocorals show a similar distribution to *Errina* spp., although suitability is less pronounced in southern regions (Figure 3-13). Habitat suitability is predicted to very slightly improve in southern areas in the future. High precision is associated with high suitability in many areas, and overall is slightly greater around future predictions than historic.

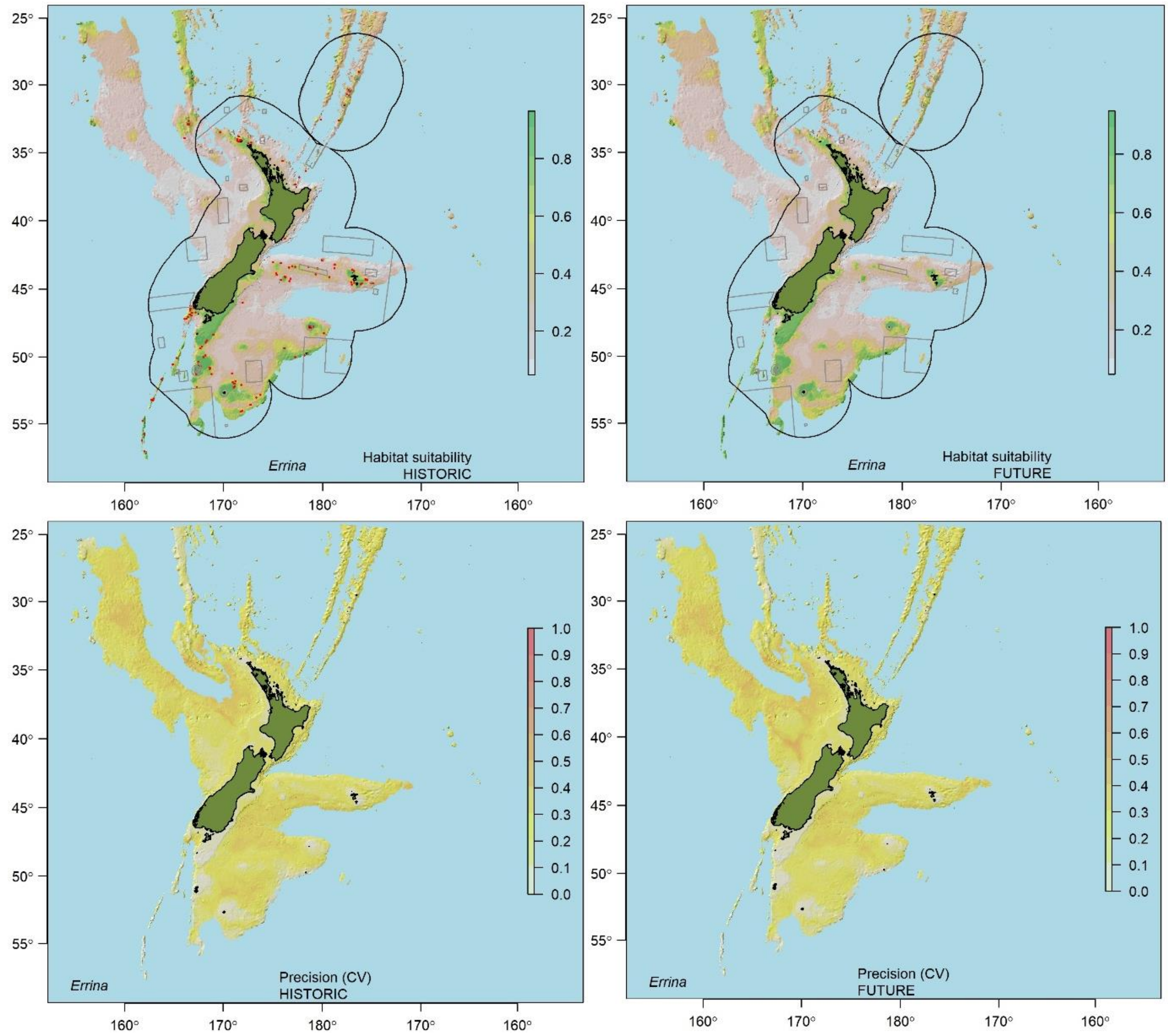


Figure 3-12: ***Errina*** **spp. Predicted habitat suitability.** See Figure 3-2 caption for further details.

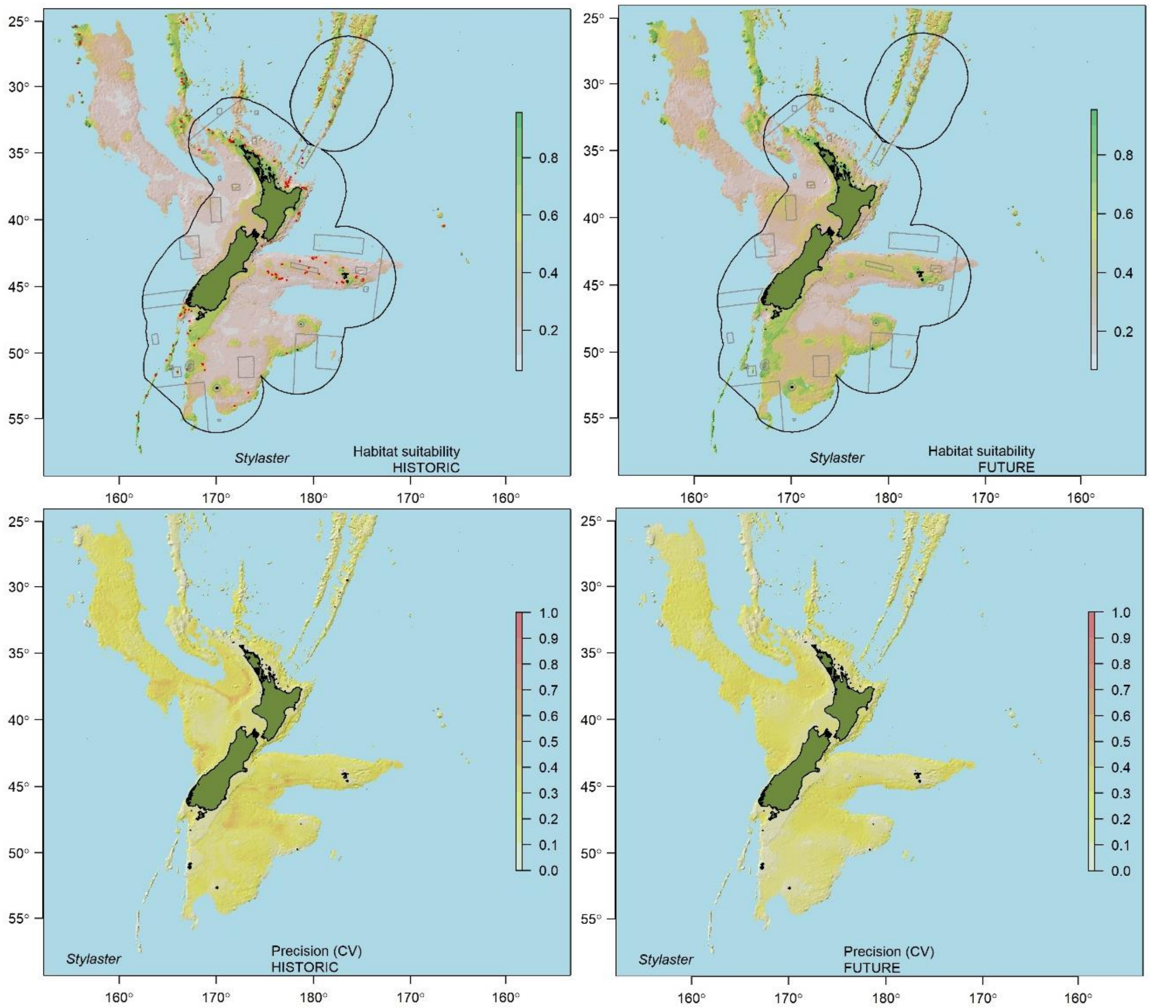


Figure 3-13: *Stylaster* spp. Predicted habitat suitability.See Figure 3-2 caption for further details.

3.5 Overlap with trawl fishing footprint

Spatial representation of the combined aggregated area swept by bottom trawling for the fishing years from 2007–08 to 2017–18 shows that the highest fishing intensity in recent years has focused on several main areas: the Chatham Rise, especially west of the Chatham Islands but excluding a shallow central region lying within a BPA; a contiguous region of the Stewart-Snares shelf and western Campbell Plateau extending south to below the Auckland Islands; the perimeter of the Campbell Rise, especially the eastern slopes, and perimeters of the Pukaki Rise and Bounty Plateau; the central west coast of mainland New Zealand, especially off the northern South Island; an area of the Challenger Plateau bordering the EEZ; and the shelf/slope along much of the eastern coastline of New Zealand (Figure 3-14).

The overlap of intense bottom trawling effort with the predicted distribution of *Enallopsammia rostrata* is mainly restricted to a few locations around the deeper parts of the eastern Chatham Rise, the sub-Antarctic Plateau, Puysegur Shelf, and the central east coast of the North Island (Figure 3-15); this overlap increases slightly for the predicted future distribution of this species, notably off the South Island west coast (Figure 3-16). The overlap is less for *Solenosmilia variabilis* and is limited to a few spots on the Chatham Rise, the southern plateaus, and the North Island east coast; the overlap is slightly greater for the predicted future habitat suitability, notably in a small region northeast of the Auckland Islands. The shallower distribution of *Goniocorella dumosa* leads to a much greater overlap with bottom trawling than these other two branching corals. Areas of overlap exist around the perimeter of the Chatham Rise, the Stewart-Snares and Auckland Islands shelfs, and around much of mainland New Zealand; the pattern of overlap changes little for the predicted future distribution. Bottom trawling overlaps with *Madrepora oculata* distribution mainly off the central east coast of the South Island, and around much of the North Island; the total number of overlapping cells isn't predicted to change much in the future, but an increase is suggested off the South Island west coast.

For the alcyonaceans modelled, the similar habitat suitability of *Paragorgia* spp., *Primnoa* spp., and *Keratoisis* & *Lepidisis* spp. translates to similar overlaps with bottom trawling, with just a few small regions of overlap along the Chatham Rise and in the sub-Antarctic. In the future, the overlaps for *Paragorgia* spp. and *Primnoa* spp. are predicted to be more focussed around the northern slopes of the Auckland Islands shelf, and for *Primnoa* spp. also across parts of the central Chatham Rise. For *Keratoisis* & *Lepidisis* spp. future overlap is predicted to be greatest off the South Island west coast. For *Corallium* spp. the overlap between trawling and habitat is relatively low – both for the current and future predicted suitability for this genus – and focussed on the Puysegur Bank and a few spots around the sub-Antarctic and east of the Chathams.

For black corals, fishing overlap with predicted habitat suitability for *Bathypathes* spp. is seen off the North Island east coast, east of the Chathams, the Puysegur Bank and a few spots around the sub-Antarctic; this is not predicted to change appreciably over time. For *Leiopathes* spp., overlaps with fishing are mainly seen on the northern Chatham Rise and around the North Island; in the future overlap is predicted to increase off the South Island west coast and decrease on the Chatham Rise.

For the anthoathecatan genera *Errina* and *Stylaster*, which show similar distributions, overlap with trawling is evident around the Stewart-Snares Shelf, Auckland Islands shelf, South Island east coast, the Chatham Rise, and around much of the North Island (particularly, for *Stylaster* spp., on the east coast). These areas of overlap decrease over time, especially along the North Island east coast, and especially for *Stylaster* spp.

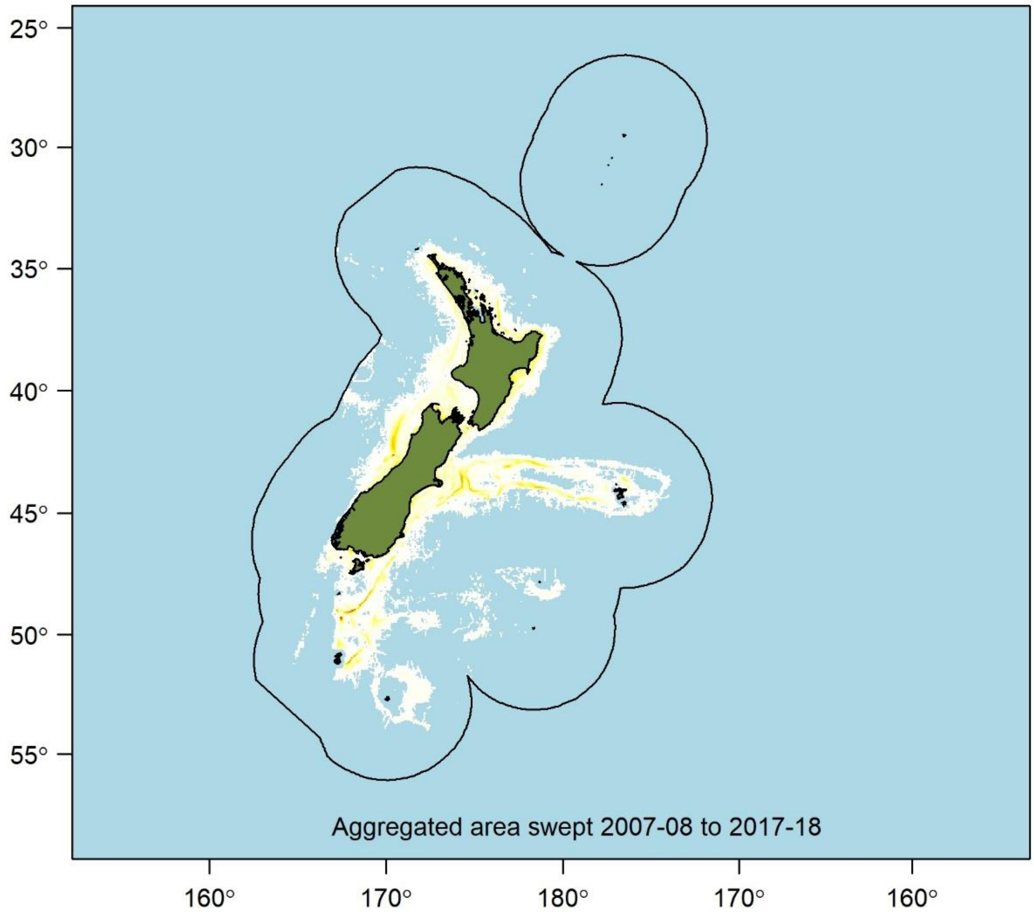


Figure 3-14: Fishing footprint. Aggregated area swept by bottom trawling for inshore and deepwater target species for the fishing years 2007–08 to 2017–18.

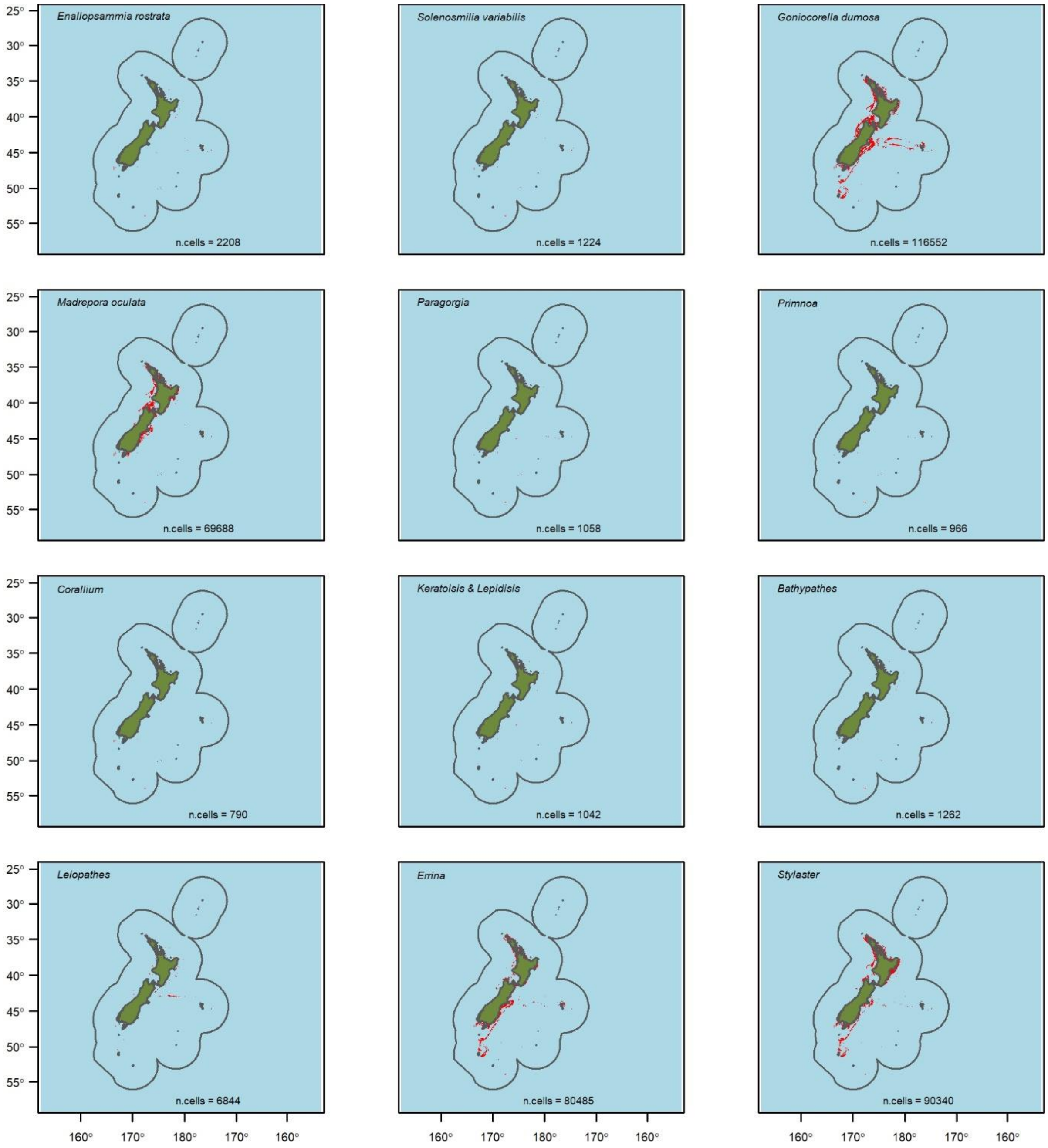


Figure 3-15: Overlap of bottom trawl fishing effort with HISTORICAL habitat suitability. Red coloured cells signify areas with both the greatest 25% aggregated swept area (inshore and deepwater target fisheries combined, 2007–08 to 2017–18) and the greatest 25% habitat suitability. N.cells = the number of overlapping 1x1 km cells.

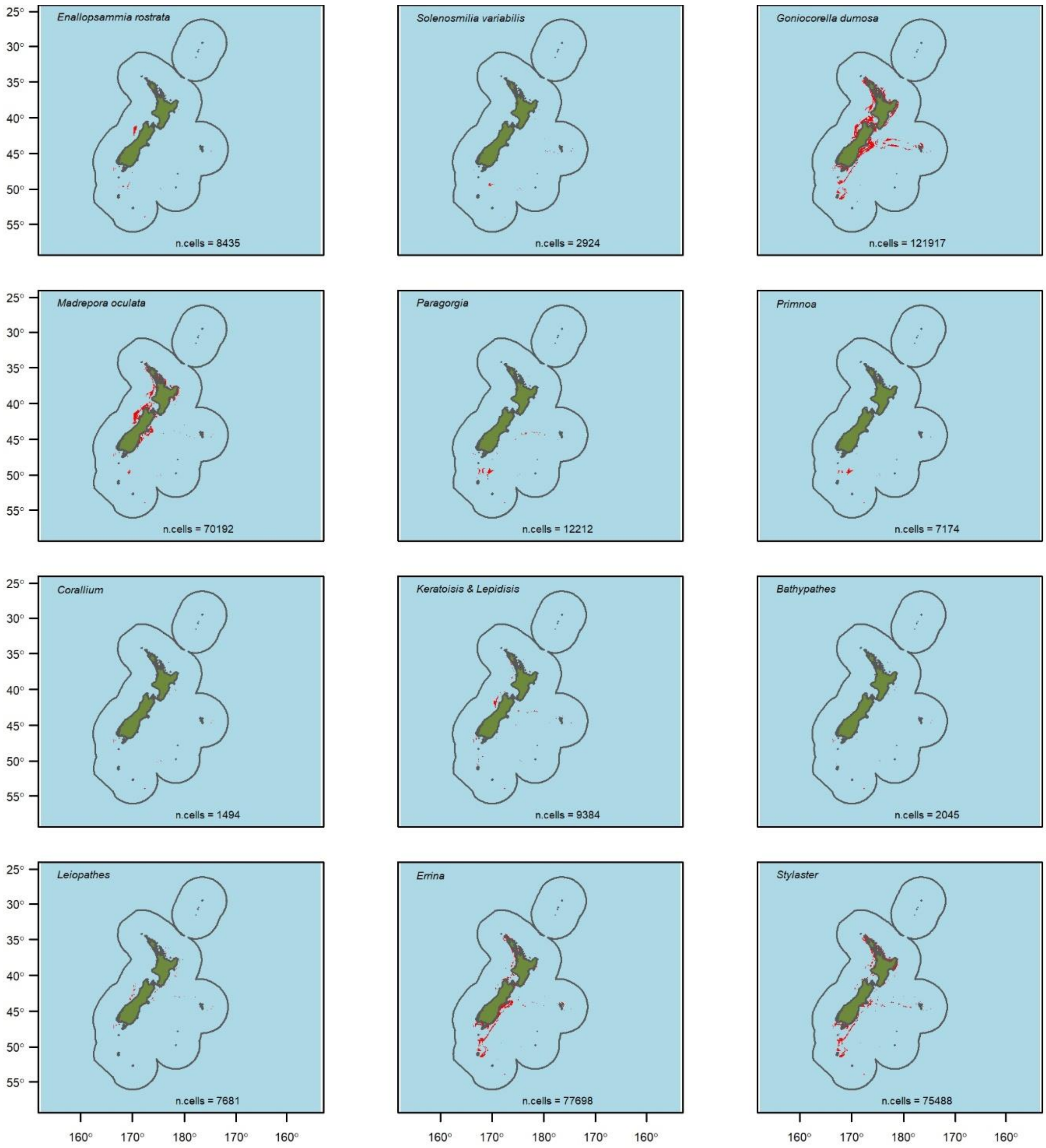


Figure 3-16: Overlap of bottom trawl fishing effort with FUTURE habitat suitability. Red coloured cells signify areas with both the greatest 25% aggregated swept area (inshore and deepwater target fisheries combined, 2007–08 to 2017–18) and the greatest 25% habitat suitability. N.cells = the number of overlapping 1x1 km cells.

4 Summary and discussion

This analysis represents the first use of the New Zealand Earth System Model to predict the future suitability of seafloor environmental conditions for supporting protected corals in the region. Forecasts from the ESM tuned to local oceanic and atmospheric conditions, along with substantial additions to species occurrence records and habitat suitability model development in recent years, has made it possible to provide some advancements in predictive models for several protected coral taxa, and produce models for two taxa for the first time.

Habitat suitability for the four species of scleractinian branching corals modelled is predicted to remain similar or even improve in some areas at the end of the century, especially for *Enallopsammia rostrata* and *Solenosmilia variabilis*. Models for these two species (and to a lesser extent the other two scleractinian genera modelled) are strongly driven by *Aragonite* concentration (see Appendix B and Appendix C). Concentration of *Aragonite* is positively correlated with prevalence for these species for values between slightly undersaturated and slightly oversaturated, but for higher concentrations the correlation becomes negative. This suggests that once seawater is fully saturated with *Aragonite* other factors may come into play to limit habitat suitability, with *Aragonite* potentially acting as a proxy for other, unknown, parameters. Other key environmental variables in the models for these species, *Oxygen* and *Nitrogen* in particular, have previously been shown to be important drivers of habitat suitability for them (e.g. Anderson et al. 2015).

Just as *Aragonite* is important for scleractinians, the *Calcite* polymorph of $CaCO_3$ is a key predictor in the models for alcyonacean corals, especially *Paragorgia* spp., for which also a peak in habitat suitability occurs at a level above full saturation, followed by decreasing suitability at higher concentrations. Benthic *Nitrogen* is also an important predictor for the alcyonacean taxa modelled, overall showing a negative correlation with habitat suitability.

For the two black coral genera modelled, dissolved oxygen was the key driver of habitat suitability. Peaks in suitability occur near the lower end of the range of values experienced by the sampled populations, with higher values of oxygen not improving suitability; again, this may indicate that other factors not accounted for in the models may be coming into play with these species.

Habitat suitability for the two anthoathecatan taxa is driven in the models mainly by fixed parameters (*Mud* and *Depth* for *Errina* spp.; *BPI-fine*, *Mud*, and *Depth* for *Stylaster* spp.). Because of this, changes in predicted habitat suitability over time for these species are very minor. Although the other taxa modelled generally showed more change in habitat suitability over time than these anthoathecatans, this was less obvious for *Corallium* and the two black coral genera as the non-fixed variables that models for these species were driven by (primarily benthic *Nitrogen* and dissolved *Oxygen*) were not predicted to change greatly over time (Appendix A).

Model accuracy was not tested with fully independent data, but the cross-validation procedure provided a useful test of model reliability by setting aside portions of the input data for comparison with model predictions. The AUC values generated by this method indicated more than adequate performance (>0.7) for all models and excellent performance (>0.8) for 18 of the 24 individual models; an AUC of over 0.9 was achieved in both model for one taxon, *Corallium* spp.. Rates of correct identification of presence (sensitivity) and absence (specificity) were mostly high, falling below 0.7 in only one case (*Goniocorella dumosa*). Model prediction bias for most taxa, and both models, tended to be in the direction of high sensitivity/low specificity. In this aspect the two model types sometimes differed, with specificity higher in one and sensitivity higher in the other for

Solenosmilia spp., *Paragorgia* spp., and *Keratoisis* & *Leiopathes* spp.. The calculation of cut-off values may be useful when considering these prediction grids in spatial planning exercises – providing values for each taxon that constitute a "valuable" cell.

There remains a level of uncertainty in model predictions, for several reasons. A key reason is that ESM simulations are based on a spatial resolution that is very coarse (0.2° or 1°) compared to the scale at which environmental conditions are known to affect benthic marine invertebrates, although this was partially ameliorated by upscaling predictor variables to a 1x1 km resolution where possible. Grids of some environmental data are available from other sources at a higher resolution, but a trade-off is necessary when predicting to future conditions, as current and future data layers need to have matching provenance to be sure that changing conditions are the sole driver of any changing habitat suitability over time. Uncertainty in predictions may also come from the influence of any important explanatory variables not available to the models, from variability in catchability among the different sampling methods used, errors in recording or specimen identification, and model inadequacies.

Model predictions are generally in good agreement with earlier models for the same taxa. Predicted distributions for *Enallopsammia rostrata* are similar to those of Anderson et al. (2015) and Anderson et al. (2014), which also showed high habitat suitability in northern regions, and similar patterns around the Chatham Rise and sub-Antarctic. As in the current study, future predictions from Anderson et al. (2015) also showed some improved habitat suitability on the Chatham Rise and less change in the North, but overall predicted lower future suitability for the species. The distribution predicted for this species by Georgian et al. (2019) highlighted the hotspots and these, showing highest suitability on the southeast Chatham Rise, southern Bounty Plateau, and the ridge systems to the north and south of New Zealand, are in general agreement with our study.

Both this study and that of Anderson et al. (2015) show increasing future habitat suitability for *Solenosmilia variabilis* in the easternmost Chatham Rise and little change in the north. The biggest difference between models is in the increasing habitat suitability predicted for shallow areas in the current study. The predictions of Anderson et al. (2014) show some differences, with higher habitat suitability in parts of the sub-Antarctic and Challenger Plateau not seen in the more recent studies. The models of Georgian et al. (2019) show broad agreement with the current distribution predicted in this study, although they lack any evidence of higher suitability in the easternmost Chatham Rise or northern Lord Howe Rise.

Because of the shallower depth cut-off in our models, predictions of high habitat suitability for *Goniocorella dumosa* are shown for some near-coast areas where predictions were previously unavailable. In other areas the new models gave very similar results to Anderson et al. (2015) and Anderson et al. (2014), with focus of predicted distribution on the shallow crest region of the Chatham Rise and a few patches around the sub-Antarctic. For future predictions, Anderson et al. (2015) indicated improved habitat suitability on the Chatham Rise and sub-Antarctic but unchanged elsewhere; our models show no appreciable change in any location. The Georgian et al. 2019 models agree well with our models but, with their deeper maximum model extent, show some additional areas of suitable habitat in deeper water to the northwest of New Zealand.

The new models show some areas with high habitat suitability for *Madrepora oculata* on the Lord Howe Rise and West Norfolk Ridge that Anderson et al. (2015) did not. This may be due to the influence of several new presence records for the species in that area, otherwise predicted distributions were mostly similar between these studies and that of Anderson et al. (2014). Changes in habitat suitability over time predicted by Anderson et al. (2015) were minor in northern regions but lower everywhere, whereas suitability was predicted to slightly improve in a few areas in the new study. Predictions are also consistent with those of Georgian et al. (2019), although that study indicated a more definitive band of high suitability around the slopes of the Chatham Rise, and additional patches of higher suitability on the western Challenger Plateau.

A substantial increase in habitat suitability for *Paragorgia* spp. was predicted by Anderson et al. (2015) but this was not confirmed by our study. Otherwise, the key areas for this genus were predicted by both studies to be around the Campbell and Bounty Plateaus. The Anderson et al. (2014) predictions were similar, although that study indicated some highly suitable areas on the sub-Antarctic that were not predicted by the later studies.

Several new records of *Primnoa* spp. in northern regions led to some areas of high habitat suitability on the Lord Howe Rise and West Norfolk Ridge that were not identified by Anderson et al (2015). Otherwise predictions were very similar to those of both Anderson et al. (2015) and Anderson et al. (2014) where most highly suitable habitat is predicted to be south of the Chatham Rise. Changes in habitat suitability in the future were not predicted to be strong by either study.

The predicted distribution of *Keratoisis* & *Lepidisis* spp. was very similar to that shown by Anderson et al. (2015) but some shallower areas on the Chatham Rise and Challenger Plateau were predicted by Anderson et al. (2014) to have high habitat suitability; and these were not seen in newer models. Both this study and Anderson et al. (2015) indicate little change over time in northern regions, but improved habitat suitability on the Chatham Rise (especially along the crest) is a prediction common to both models.

Some new presence records in the north again have led to higher habitat suitability for the two black coral genera in some areas (especially for *Leiopathes* spp.), but otherwise the predictions of Anderson et al. (2015) and Anderson et al. (2014) are very similar to ours. Our predictions correspond to those of Anderson et al. (2015) in predicting little change over time in any area, although notably both models show an increase in habitat suitability in the southern Challenger Plateau.

Further improvements in habitat suitability models for predicting the current distribution of protected corals will be possible with the continued expansion of the area around New Zealand surveyed by towed camera and multibeam echosounders. The high data resolution obtained from such surveys (~25x25 m) will allow model development on a scale at which environmental conditions, terrain and sediment types in particular, are likely to be most strongly correlated with settlement, growth, and survival of benthic marine invertebrates. Substantial improvements in prediction of the future distribution of corals will however require advancements in the precision of ESM model outputs, with resolution much finer than the 1° or 0.2° currently available.

There is considerable variability among coral taxa in the overlap of predicted distributions with bottom trawling. The taxa most vulnerable are the thicket-forming corals *Goniocorella dumosa* and *Madrepora oculata*, and the hydrocorals *Errina* spp. and *Stylaster* spp. These taxa have distributions extending into shallower depths, and into key fishery areas on the South Island northern west coast and central east coast, the Chatham Rise, and western regions of the sub-Antarctic.

5 Acknowledgements

We thank Brent Wood (NIWA) for GIS support and, along with Di Tracey, Sadie Mills, David Fisher, Diana MacPherson, and Ashley Rowden (all NIWA), for the compilation of new species presence records. Thanks also to Karen Lisa Tunley and Riki Mules (FNZ) for supply of the trawl footprint data layers. Data compilation was carried out partly under a SPRFMO contract (SPR201801) for which we are thankful to MPI. We acknowledge the support of Shannon Weaver (DOC) and funding through DOC contract POP2018-01.

6 References

Allouche, O.; Tsoar, A; Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43, 1223–1232. doi: 10.1111/j.1365-2664.2006.01214.x

Anderson, O.; Tracey, D.; Bostock, H.; Williams, M.; Clark, M. (2014). Refined habitat suitability modelling for protected coral species in the New Zealand EEZ. NIWA Client Report prepared for Department of Conservation. WLG2014-69.

Anderson, O.; Mikaloff Fletcher, S.; Bostock, H. (2015). Development of models for predicting future distributions of protected coral species in the New Zealand Region. NIWA Client Report prepared for Department of Conservation. WLG2015-65.

Anderson OF, Guinotte JM, Rowden AA, Tracey DM, Mackay KA, Clark MR. 2016b. Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand. 2016b. *Deep Sea Research Part I: Oceanographic Research Papers 115*:265-92.

Anderson, O.F.; Ballara, S.L.; Edwards, C.T.T. (2017). Fish and invertebrate bycatch and discards in New Zealand orange roughy and oreo trawl fisheries from 2001–02 until 2014–15. *New Zealand Aquatic Environment and Biodiversity Report No. 190*. 216 p.

Baird, S.J.; Mules, R. (2019). Extent of bottom contact by New Zealand commercial trawl fishing for deepwater Tier 1 and Tier 2 target species determined using CatchMapper software, fishing years 2008–17. *New Zealand Aquatic Environment and Biodiversity Report No*. 229. 106 p.

Barbet-Massin, M.; Jiguet, F.; Albert, C.H.; Thuiller, W. (2012). Selecting pseudoabsences for species distribution models: how, where and how many. *Methods Ecol. Evol*. 3, 327–338.

Behrens, E.; Williams, J.; Morgenstern, O.; Sutton, P.; Rickard, G.; Williams, M. (2019). Local Grid Refinement in New Zealand's Earth System Model: Tasman Sea Ocean Circulation Improvements and Super-Gyre Circulation Implications. *Earth and Space Science Open Archive*.

Bostock, H.C., Tracey, D.M., Currie, K.I., Dunbar, G.B., Handler, M.R., Mikaloff-Fletcher, S.E., Smith, A.M., Williams, M.J.M. (2015). The carbonate mineralogy and distribution of habitat-forming deep-sea corals in the southwest pacific region. *Deep-Sea Res. I* 100, 88–104.

Bostock, B.; Jenkins, C.; Mackay, K.; Carter, L.; Nodder, S.; Orpin, A.; Pallentin, A.; Wysoczanski, R. (2018a): Distribution of surficial sediments in the ocean around New Zealand/Aotearoa. Part A: continental slope and deep ocean. *New Zealand Journal of Geology and Geophysics*. DOI:10.1080/00288306.2018.1523198

Bostock, B.; Jenkins, C.; Mackay, K.; Carter, L.; Nodder, S.; Orpin, A.; Pallentin, A.; Wysoczanski, R. (2018b): Distribution of surficial sediments in the ocean around New Zealand/Aotearoa. Part B: continental shelf. *New Zealand Journal of Geology and Geophysics*. DOI:10.1080/00288306.2018.1523199

Bowden, D.A.; Anderson, O.A.; Escobar-Flores, P.; Rowden, A.A.; Clark, M.R. (in press). Quantifying benthic biodiversity: using seafloor image data to build single-taxon and community distribution models for Chatham Rise, New Zealand. *New Zealand Aquatic Environment and Biodiversity Report No. XXX*. XX p.

Breiman L. 2001. Random forests. *Machine learning. 45(1)*:5-32.

CANZ (2008). New Zealand Region Bathymetry. NIWA Chart, Miscellaneous Series No. 85, National Institute of Water and Atmospheric Research, Wellington, New Zealand.

Compton, T.J.; Bowden, D.A.; Pitcher, R.C.; Hewitt, J.E.; Ellis, N. (2013). Biophysical patterns in benthic assemblage composition across contrasting continental margins off New Zealand. *Journal of Biogeography* 40(1): 75-89. http://dx.doi.org/10.1111/j.1365-2699.2012.02761.x

Elith, J.; C. H. Graham; R. P. Anderson; M. Dudík; S. Ferrier; A. Guisan; R. J. Hijmans; F. Huettmann; J. R. Leathwick; A. Lehmann; J. Li; L. G. Lohmann; B. A. Loiselle; G. Manion; C. Moritz; M. Nakamura; Y. Nakazawa; J. Overton; P. Townsend; S. J. Phillips; K. Richardson; R. Scachetti-Pereira; R. E. Schapire; J. Soberón; S. Williams; M. S. Wisz; N. E. Zimmermann. (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(4)*:802-13.

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

Freeman, D.J.; Marshall, B.A.; Ahyong, S.T.; Wing, S.R.; Hitchmough R.A. (2010). Conservation status of New Zealand marine invertebrates, 2009. *New Zealand Journal of Marine and Freshwater Research Vol. 44(3*). 129-148

Freeman, D.J.; Schnabel, K. Marshall, B.A.; Gordon, D.; Wing, S.R.; Tracey, D.; Hitchmough, R.A. (2013). Conservation status of New Zealand marine invertebrates, 2013. Department of Conservation Report. 25 p.

Fricko, O. et al. (2017). The marker quantification of the Shared Socioeconomic Pathway 2: A middle-of-the-road scenario for the 21st century. *Global Environmental Change 42*. 251-267. https://doi.org/10.1016/j.gloenvcha.2016.06.004

Friedman, J.H. (1991). Multivariate Adaptive Regression Splines. *Ann. Statist*. 19 (1). 1--67. doi:10.1214/aos/1176347963. https://projecteuclid.org/euclid.aos/1176347963

Georgian, S.E.; Anderson, O.F.; Rowden, A.A. (2019). Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. *Fisheries Research 211*. 256–274.

Hastie, T.; Tibshirani, R. (1990). Generalized Additive Models. London: Chapman and Hall.

Heiberger, R.; Holland, B. (2004). Statistical analysis and data display: an inter-mediate course with examples in S-PLUS R and SAS (Springer texts in statistics). Springer.

Heiberger, R.M., (2011). HH: Statistical Analysis and Data Disp.: Heiberger Holl. R. Package version 2, 2–23 http://CRAN.R-project.org/package=HH

Hijmans, R.J, Phillips, S., Leathwick, J. and Elith, J. (2011), Package 'dismo'. Available online at: http://cran.r-project.org/web/packages/dismo/index.html.

Hirzel, A.H.; Hausser, J.; Chessel, D. & Perrin, N. (2002). Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology 2002(83)*. 2027-2036

Hosmer, D.W.J.; Lemeshow, S.; Sturdivant, R.X. (2013). Applied Logistic Regression, Third Edition. John Wiley and Sons, 528 p.

Huang, Z.; Brooke, B.; Li, J. (2011) Performance of predictive models in the marine benthic environments based on predictions of sponge distribution on the Australian continental shelf. *Shelf. Ecol. Inform.*, 6 (3-4): 205-216

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. *Mar. Ecol. Prog. Ser.* 321, 267–281.

Mackay, K. A. (2007). Database documentation: SEAMOUNT. NIWA Internal Report. 42 p.

McCullagh P.; Nelder, J. A. (1989). Generalized Linear Models. London: Chapman and Hall.

Nix, H.A. (1986). A biogeographic analysis of Australian elapid snakes. Atlas of elapid snakes of Australia: Australian flora and fauna series 7 (ed. by R. Longmore), pp. 4–15. Bureau of Flora and Fauna, Canberra.

Oppel, S.; Meirinho, A.; Ramírez, I.; Gardner, B.; O'Connell, A.F.; Miller, P.I.; Louzao, M.; (2012). Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biol. Conserv*. 156, 94–104.

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*:309-17.

Phillips, S.J.; Anderson, R.P.; Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling 190(3-4)*:231-59.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reiss, H.; Birchenough, S.; Borja, A.; Buhl-Mortensen, L.; Craeymeersch, J.; Dannheim, J.; Darr, A.; Galparsoro, I.; Gogina, M.; Neumann, H.; Populus, J.; Rengstorf, A.M.; Valle, M.; van Hoey, G.; Zettler, M.L.; Degraer, S. (2015). Benthos distribution modelling and its relevance for marine ecosystem management. *Ices Journal of Marine Science 72(2)*: 297-315. http://dx.doi.org/10.1093/icesjms/fsu107

Ridgeway, G. (2019). Generalized Boosted Models: A guide to the gbm package. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=2ahUKEwijkdH3icvnAhVi7HMBHbg1B34QFjABegQIAhAB&url=https%3A%2F%2Fcran.r-project.org%2Fweb%2Fpackages%2Fgbm%2Fvignettes%2Fgbm.pdf&usg=AOvVaw3G_GwG3t2iypBBrTj4mFyt

Robert, K.; Jones, D.O.B.; Roberts, J.M.; Huvenne, V.A.I. (2016). Improving predictive mapping of deep-water habitats: Considering multiple model outputs and ensemble techniques. *Deep-Sea Research Part I-Oceanographic Research Papers 113*: 80-89. http://dx.doi.org/10.1016/j.dsr.2016.04.008

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F. and Dormann, C.F. (2017), Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40: 913-929. https://doi.org/10.1111/ecog.02881

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. *New Zealand Aquatic Environment and Biodiversity Report 27*. 49 p.

Rowden, A.A; Anderson, O.F.; Georgian, S.E.; Bowden, D.A.; Clark, M.R.; Pallentin, A.; Miller, A. (2017). High-resolution habitat suitability models for the conservation and management of vulnerable marine ecosystems on the Louisville Seamount Chain, South Pacific Ocean. *Frontiers in Marine Science 4:335*. doi: 10.3389/fmars.2017.00335

Stephenson, F; Goetz, K.; Sharp, B.R.; Mouton, T.L.; Beets, F.L.; Roberts, J.; MacDiarmid, A.B.; Constantine, R.; Lundquist, C.J. (2020). Modelling the spatial distribution of cetaceans in New Zealand waters. *Diversity and Distributions 2020;00.* 1–22. https://doi.org/10.1111/ddi.13035

Stockwell, D. R. B. (1999). Genetic algorithms II. Pages 123–144 in A. H. Fielding, editor. Machine learning methods for ecological applications. Kluwer Academic Publishers, Boston

Tracey, D.M.; Rowden, A.A.; Mackay, K.A.; Compton, T. (2011). Habitat-forming cold-water corals show affinity for seamounts in the New Zealand region. *Marine Ecology-Progress Series 430*:1-U59. http://dx.doi.org/10.3354/meps09164>

Tracey, D.; Mills S.; Macpherson, D.; Thomas, H. (2017). Identification and storage of cold-water coral bycatch specimens. Final Report prepared by NIWA for the Conservation Services Programme, Department of Conservation. INT2015-03. NIWA Client Report 2017349WN. 38 p.

https://www.doc.govt.nz/globalassets/documents/conservation/marine-and-coastal/marine-conservation-services/reports/coral-id-final-annual-report-december-2017-web.pdf

Valavi, R.; Elith, J.; Lahoz-Monfort, J.J.; Guillera-Arroita, G. (2019). blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol*. 10: 225– 232. https://doi.org/10.1111/2041-210X.13107

Vierod, A.D.T.; Guinotte, J.M.; Davies, A.J. (2014). Predicting the distribution of vulnerable marine ecosystems in the deep sea using presence-background models. *Deep-Sea Research Part II - Topical Studies in Oceanography 99*:6-18. http://dx.doi.org/10.1016/j.dsr2.2013.06.010

Williams, J.; Morgenstern, O.; Varma, V.; Behrens, E.; Hayek, W.; Oliver, H.; Dean, S.; Mullan, B.; Frame, D. (2016). Development of the New Zealand Earth System Model: NZESM. *Weather and Climate 36*:25-44.

Wood, B.; Finucci, B.; Black, J. (in prep). A standardised approach for creating spatial grids for analysing (binning) New Zealand fisheries data.

Yesson, C.; Bedford, F.; Rogers, A.D.; Taylor, M. (2015). The global distribution of deep-water Antipatharia habitat. *Deep-Sea Res. II*, http://dx.doi.org/10.1016/j.dsr2.2015.12.004i

Appendix A Environmental predictor variables

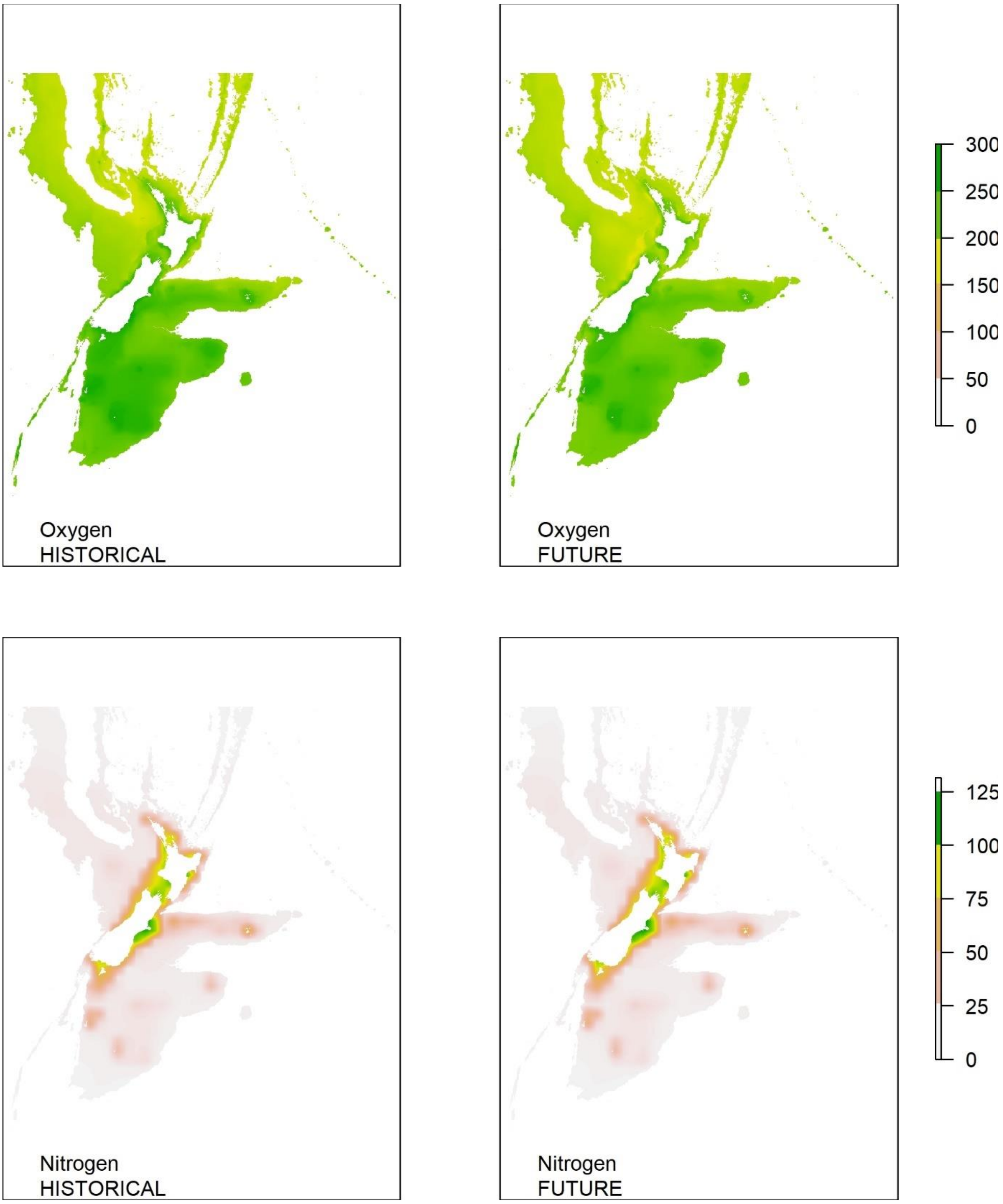


Figure A-1: Environmental variables used in the models. Top, Dissolved Oxygen (mmol/m^2); bottom, Benthic Nitrogen (mmol/m^3).

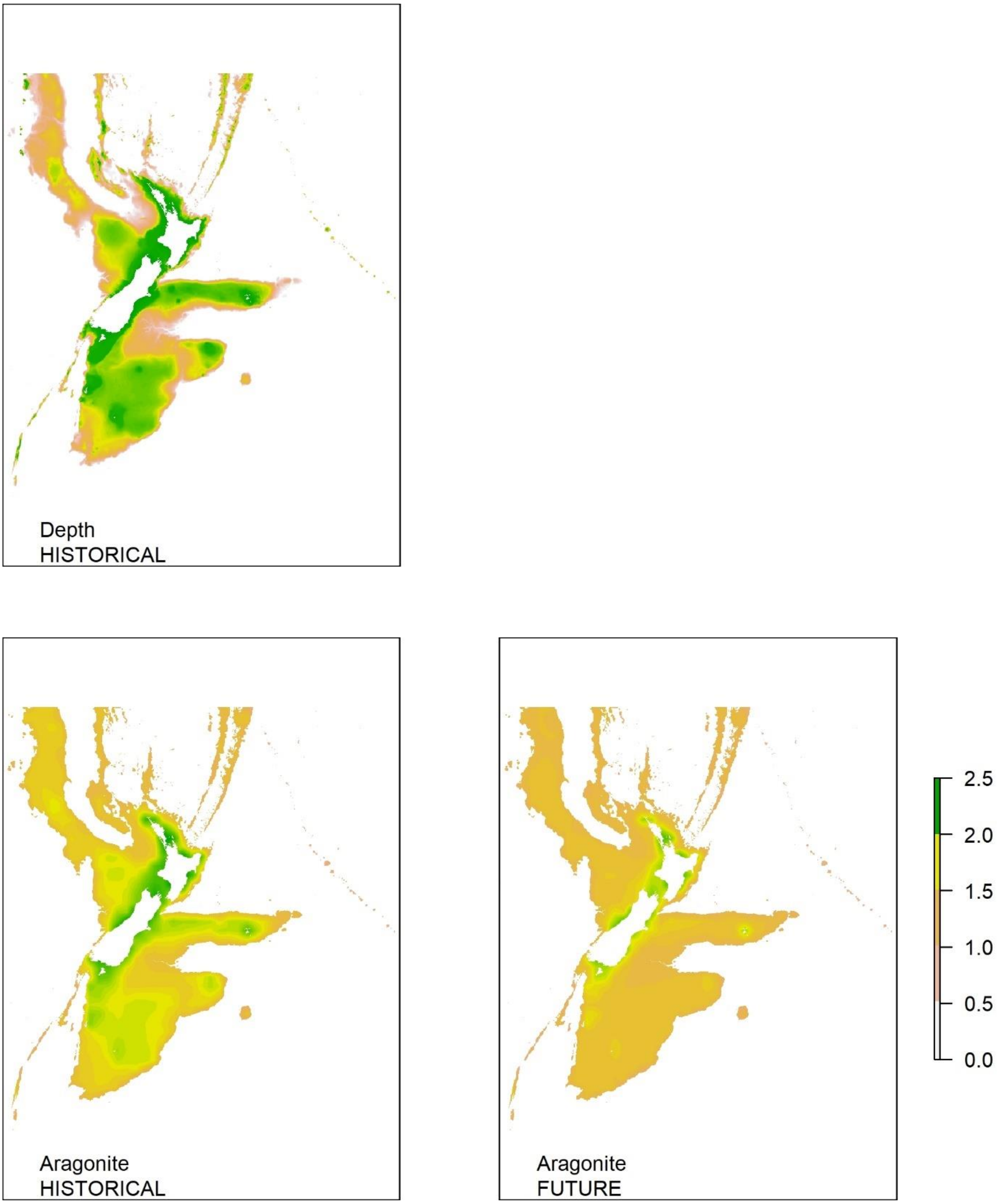


Figure A-2: Environmental variables used in the models. Top, Depth (m); bottom, Aragonite saturation.

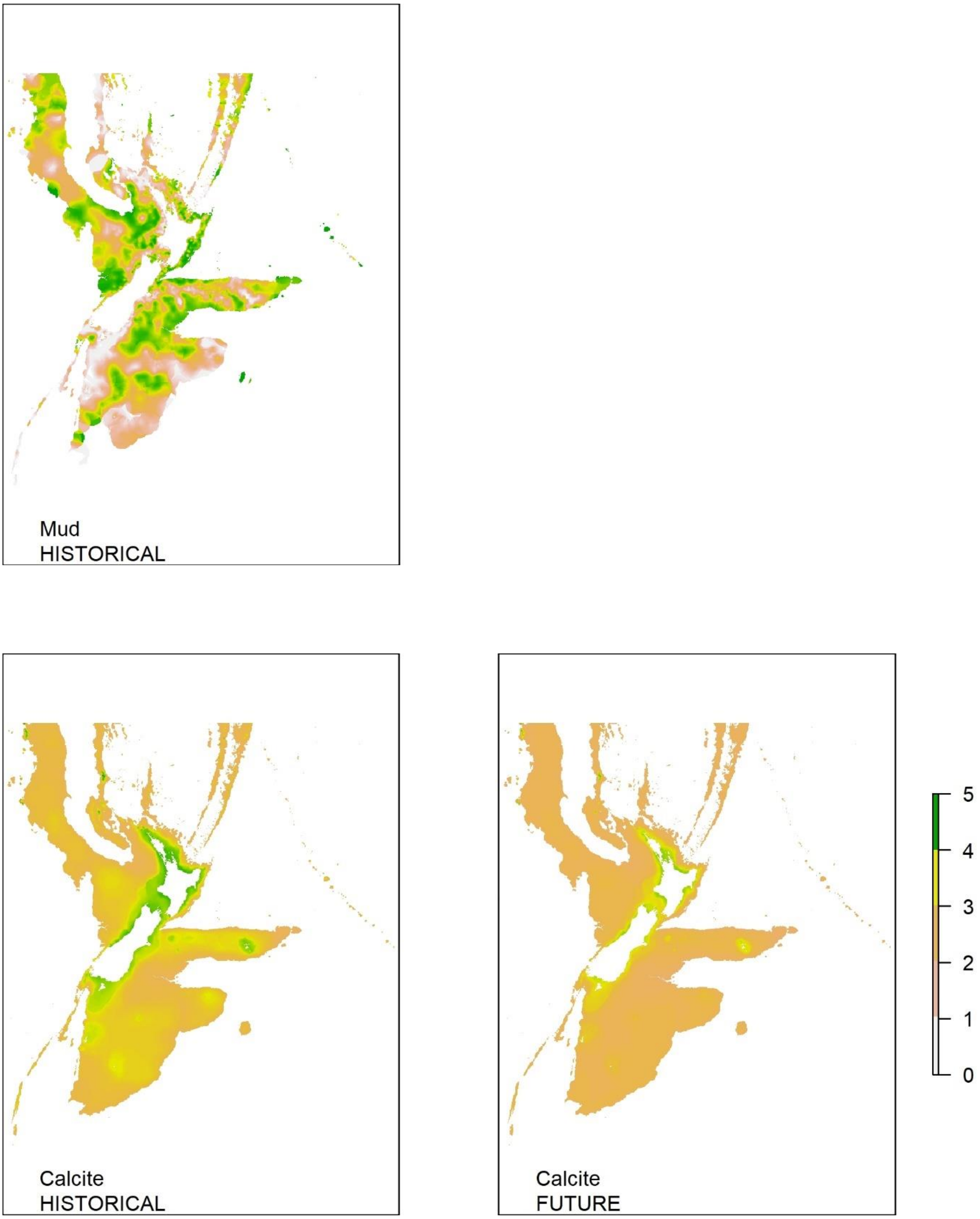
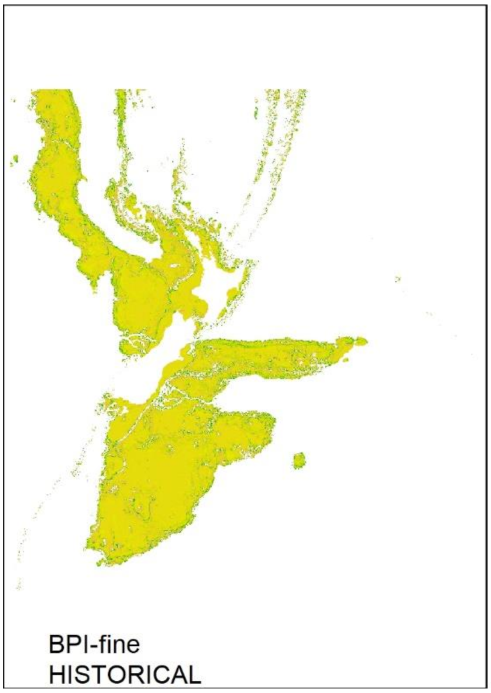


Figure A-3: Environmental variables used in the models. Top, Mud (%); bottom, Calcite saturation.



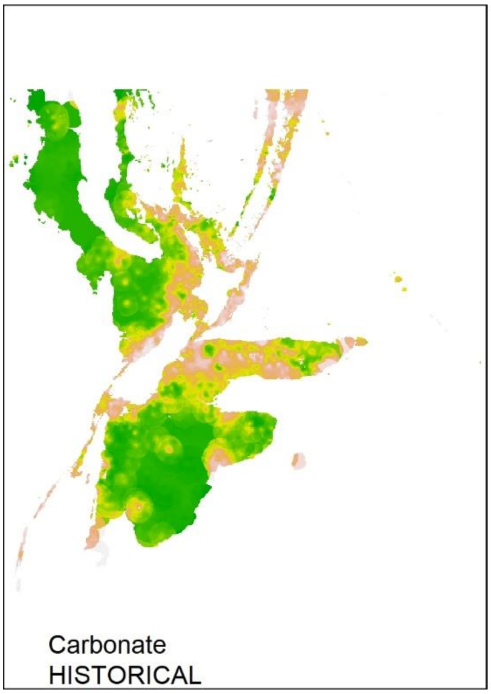


Figure A-4: Environmental variables used in the models. Top, BPI-fine; bottom, Carbonate (%).

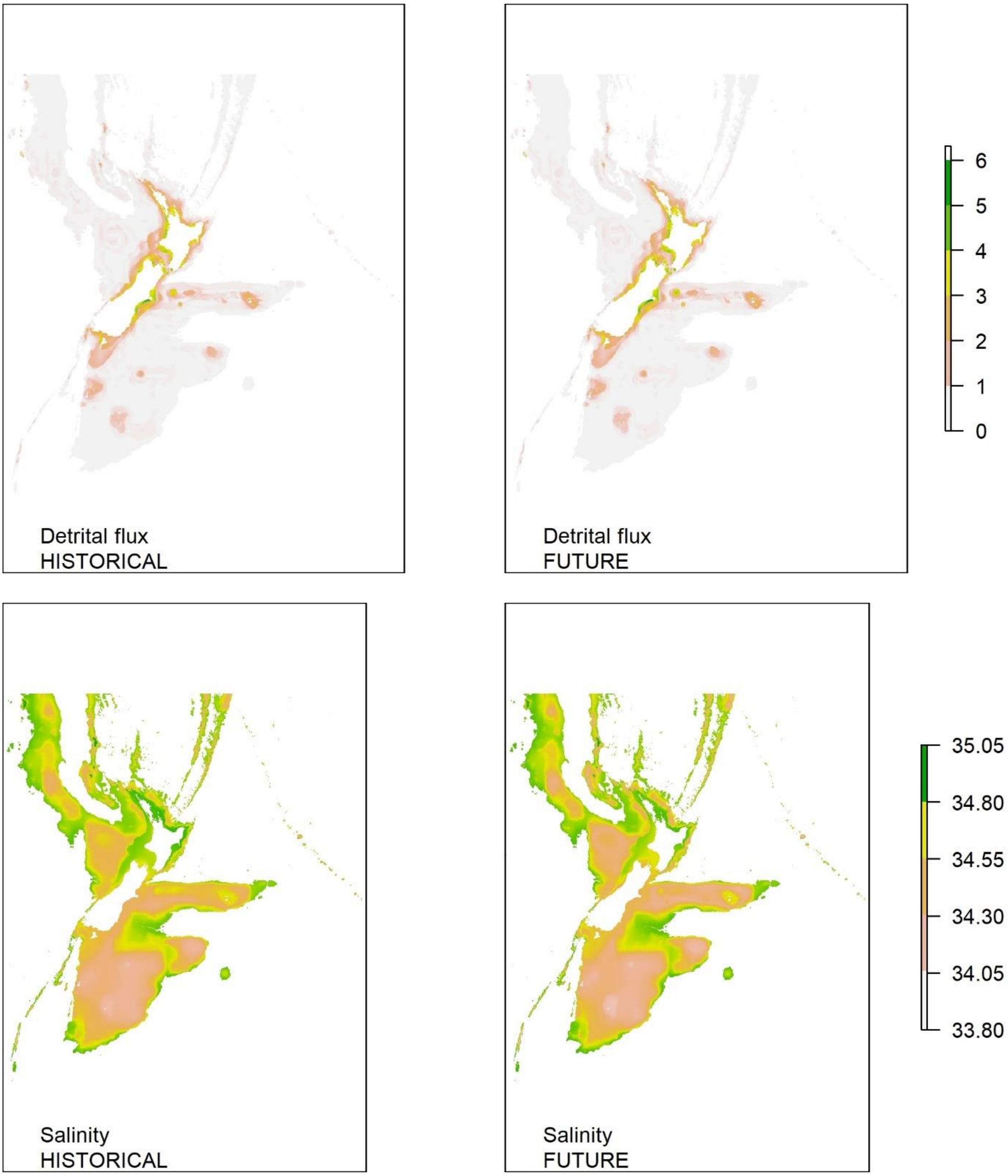


Figure A-5: Environmental variables used in the models. Top, detrital flux to seabed (mmol/m^3); bottom, Salinity (Psu).

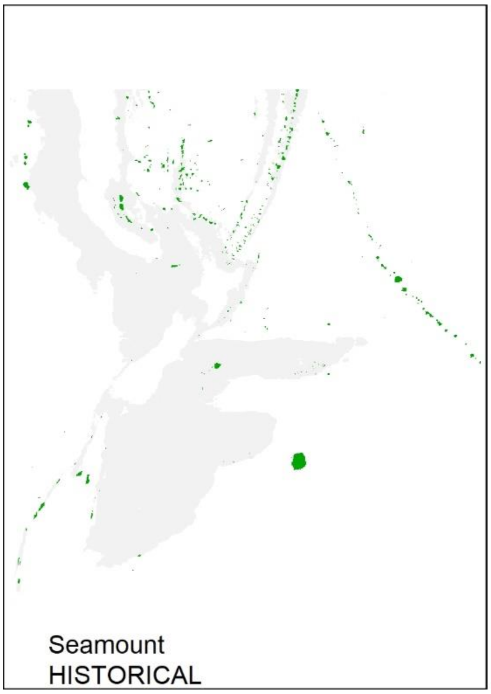


Figure A-6: Environmental variables used in the models. Top, Seamount; bottom, Slope-percent.

Appendix B Random Forest models partial dependence plots

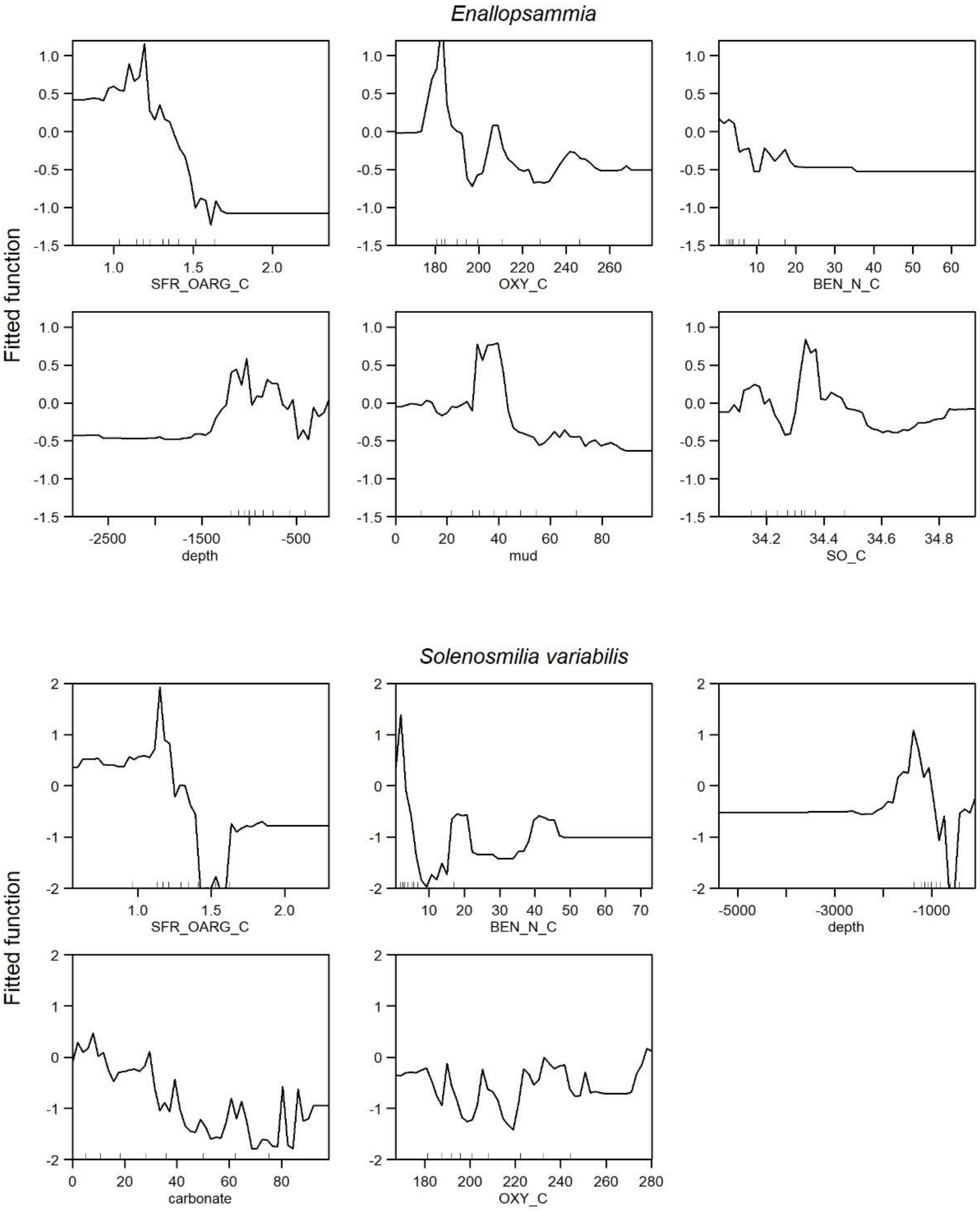
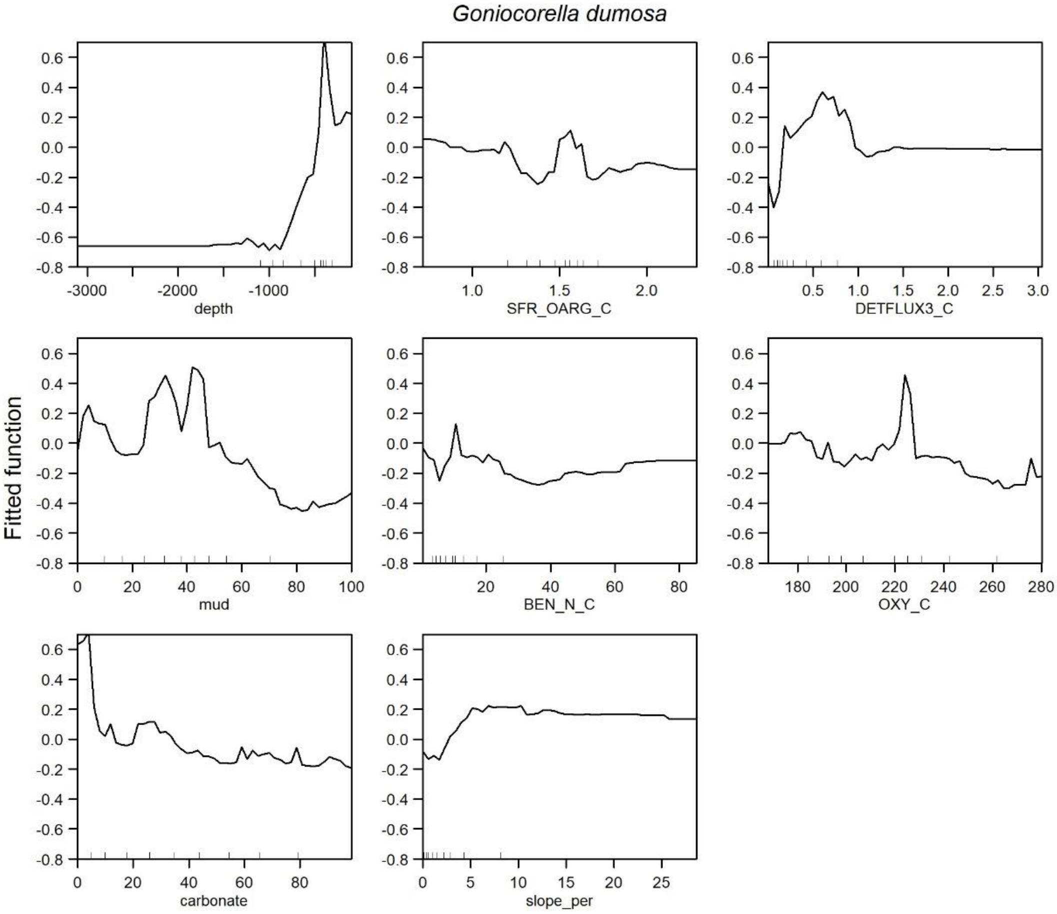


Figure B-1: Variable marginal effects in the Random Forest models for *Enallopsammia rostrata* (top) and *Solenosmilia variabilis* (bottom). See Table 2-2 for an explanation of the variables.



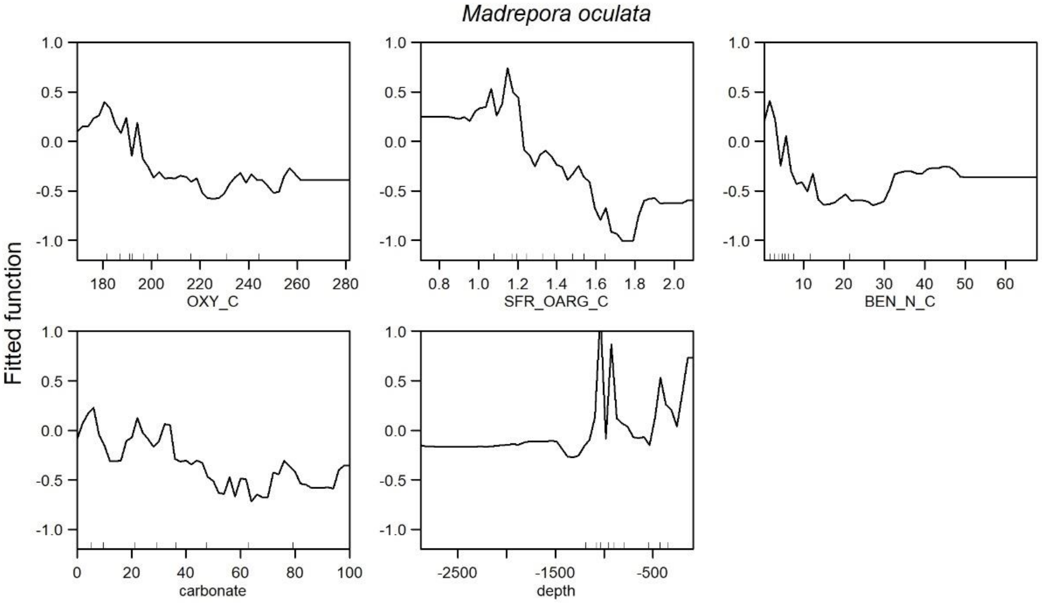
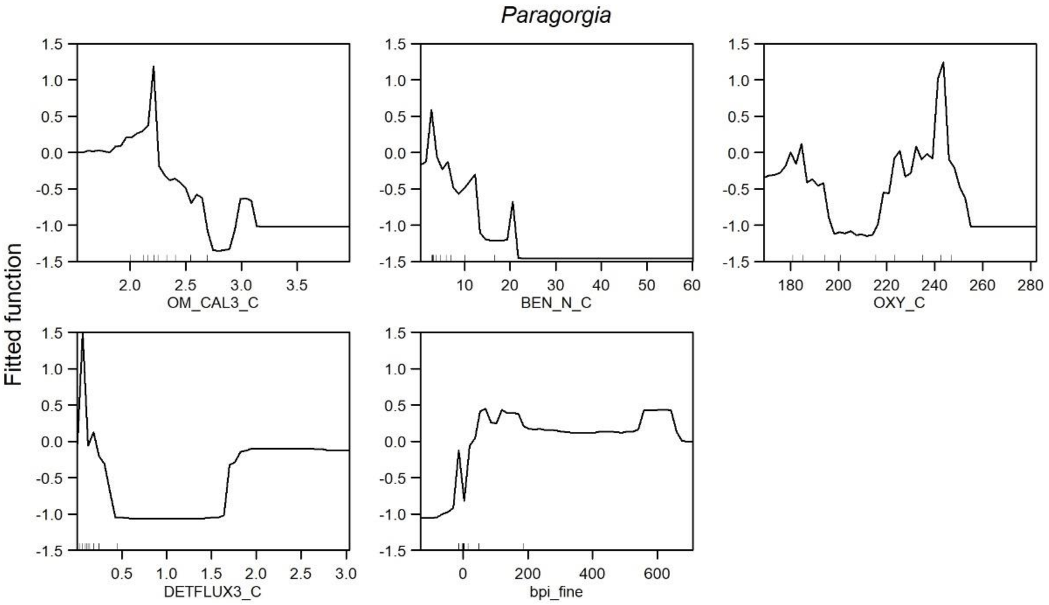


Figure B-2: Variable marginal effects in the Random Forest models for *Goniocorella dumosa* (top) and *Madrepora oculata* (bottom). See Table 2-2 for an explanation of the variables.


Paragorgia

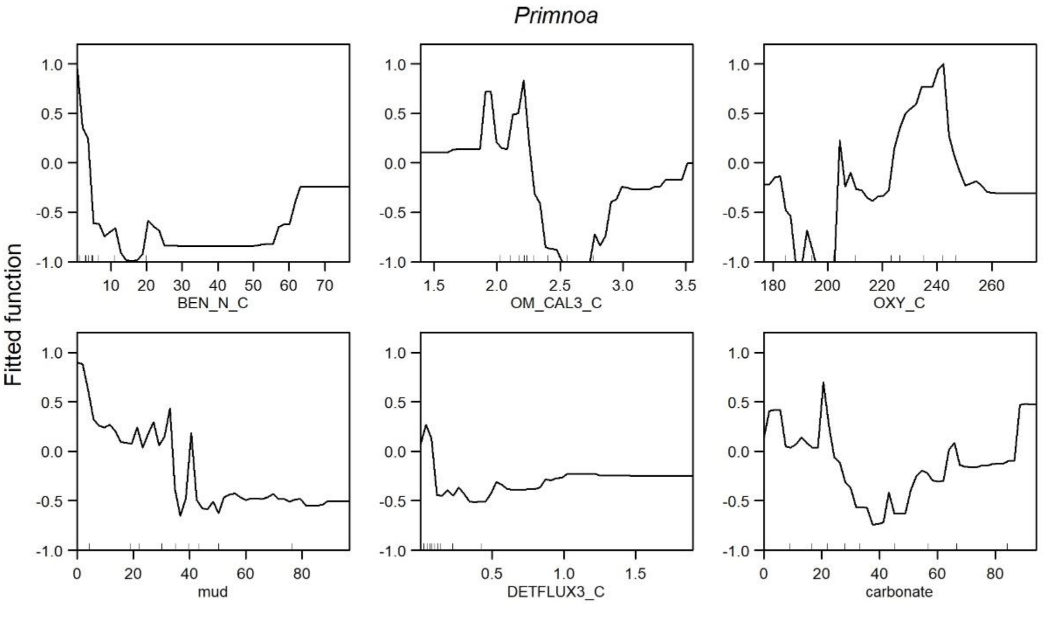

Primnoa

Figure B-3: Variable marginal effects in the Random Forest models for *Paragorgia* spp. (top) and *Primnoa* spp. (bottom). See Table 2-2 for an explanation of the variables.

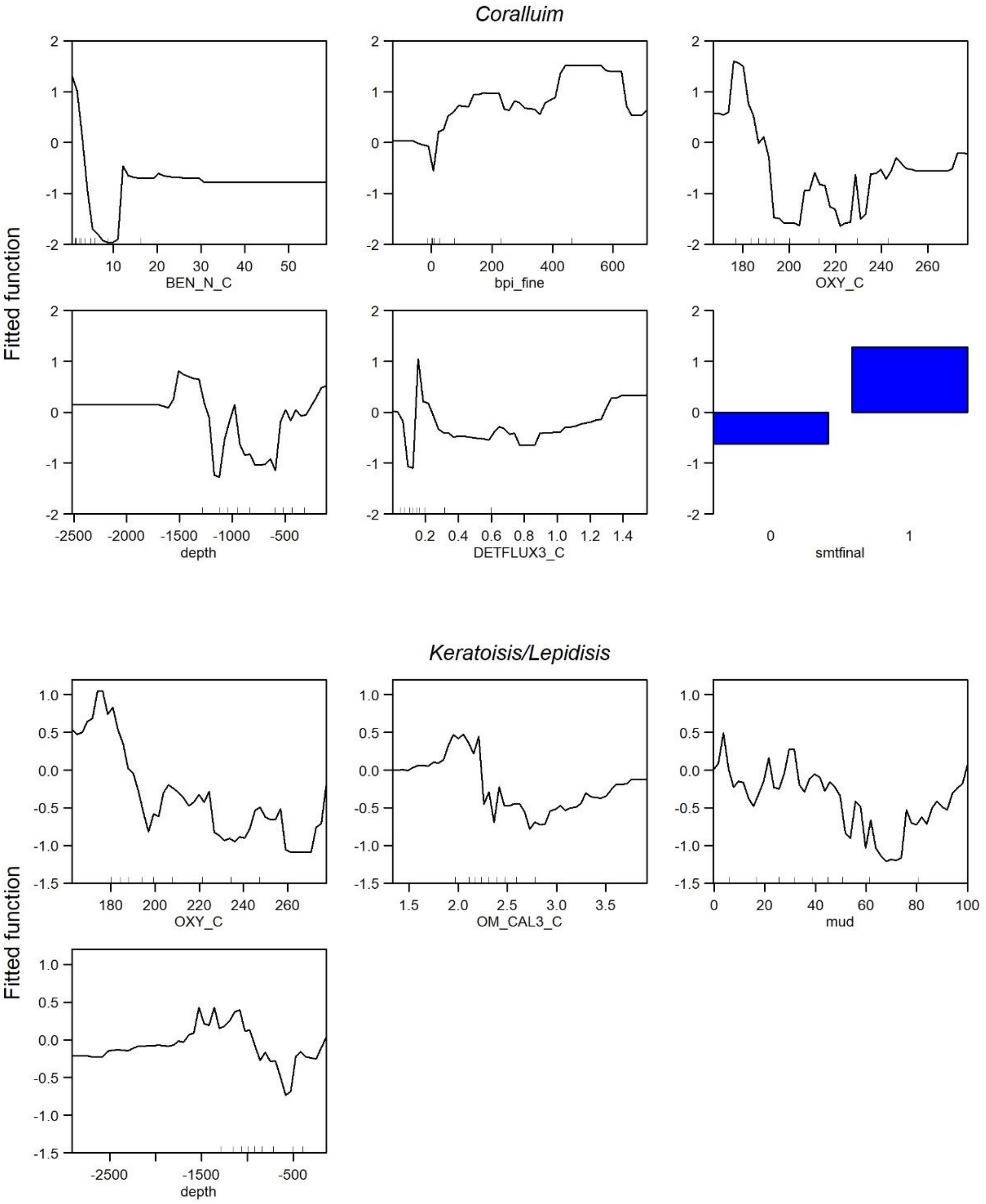


Figure B-4: Variable marginal effects in the Random Forest models for *Corallium* spp. (top) and *Keratoisis* & *Lepidisis* spp. (bottom). See Table 2-2 for an explanation of the variables.

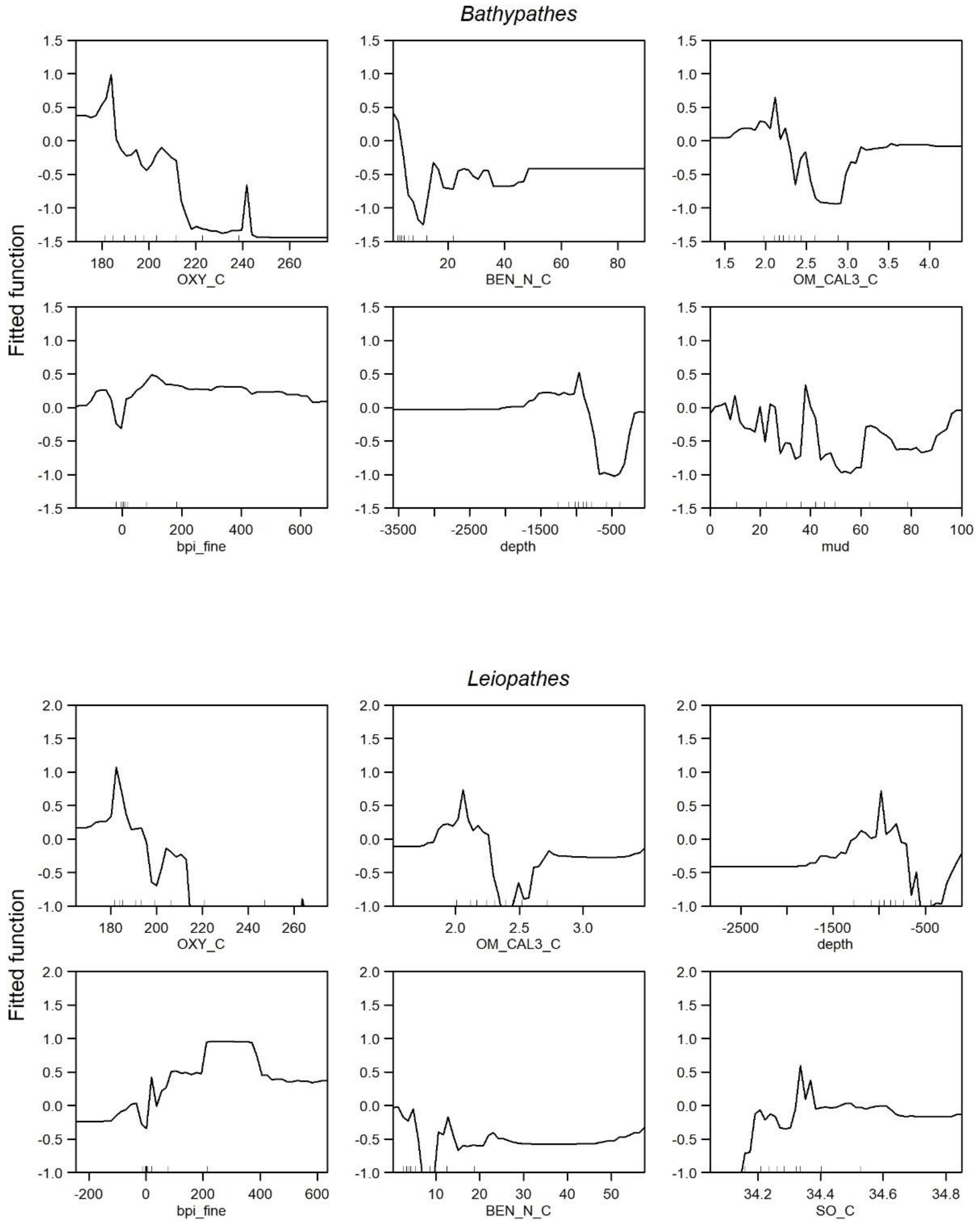


Figure B-5: Variable marginal effects in the Random Forest models for *Bathypathes* spp. (top) and *Leiopathes* spp. (bottom). See Table 2-2 for an explanation of the variables.

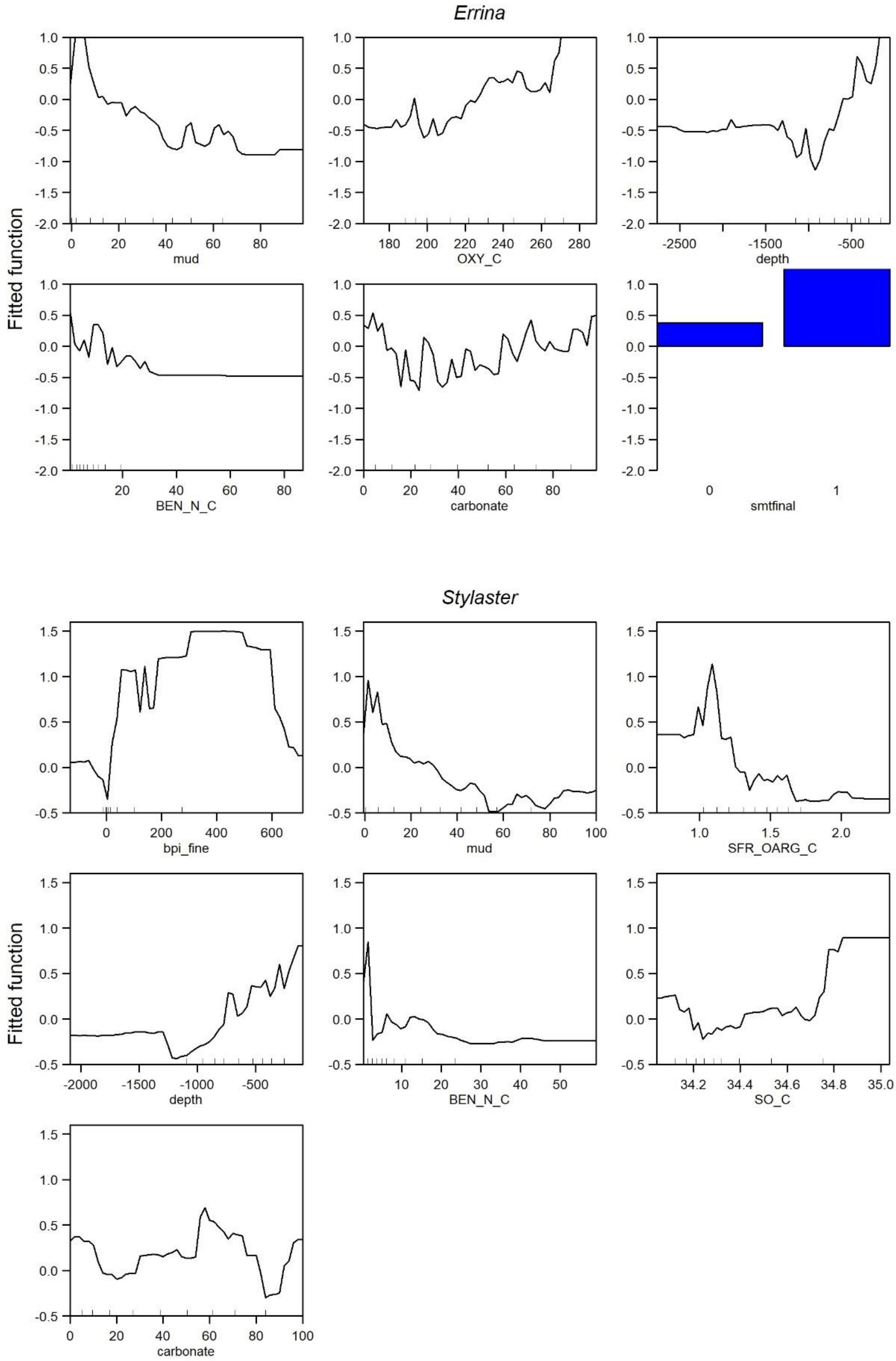
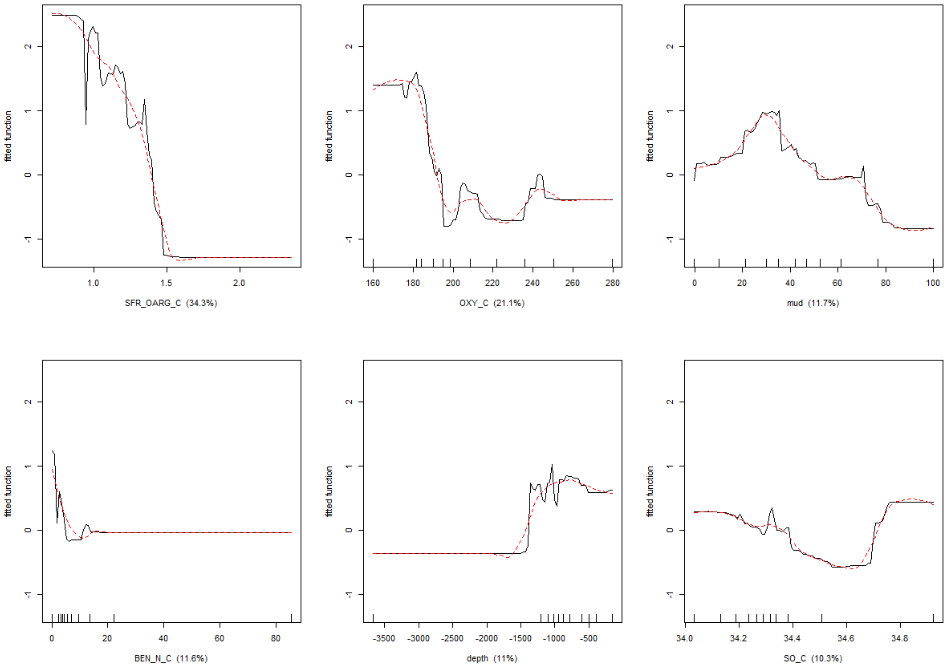


Figure B-6: Variable marginal effects in the Random Forest models for *Errina* spp. (top) and *Stylaster* spp. (bottom). See Table 2-2 for an explanation of the variables.

Appendix C Boosted Regression Tree models partial dependence plots

Enallopsammia rostrata



Solenosmilia variabilis

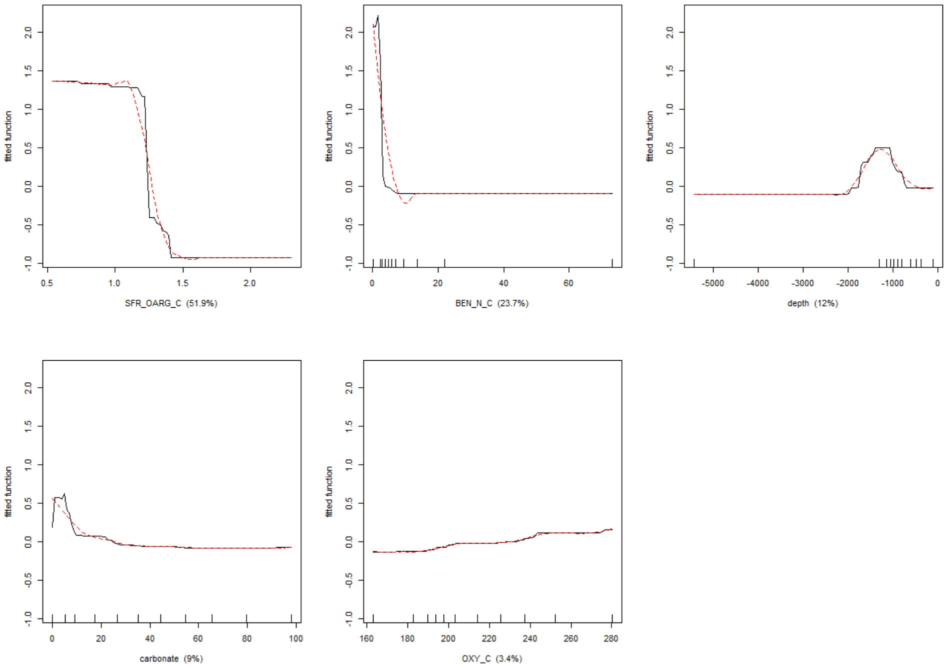
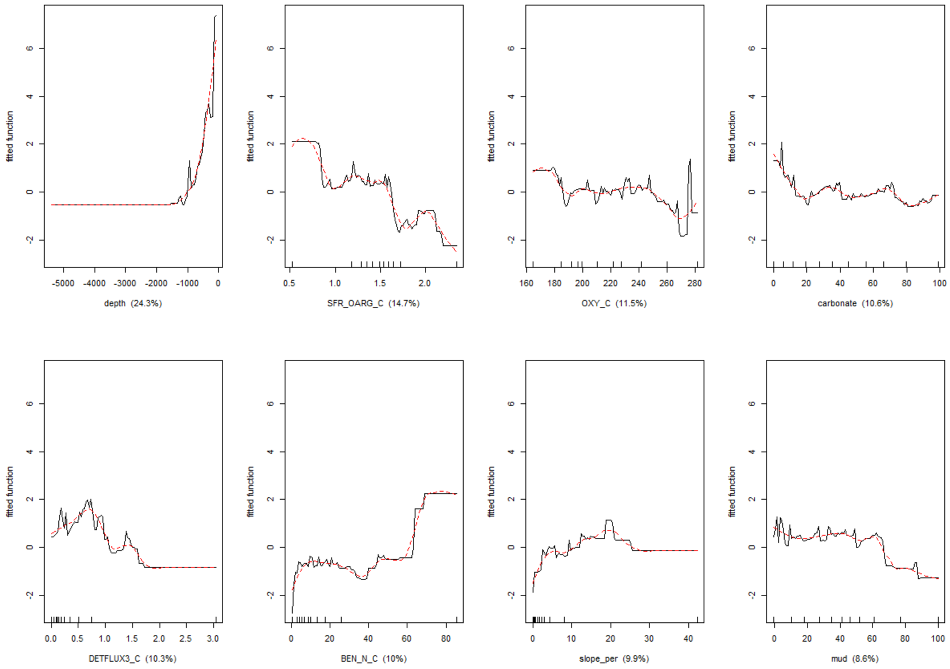


Figure C-1: Variable marginal effects in the Boosted Regression tree models for *Enallopsammia rostrata* (top) and *Solenosmilia variabilis* (bottom). See Table 2-2 for an explanation of the variables.

Goniocorella dumosa



Madrepora oculata

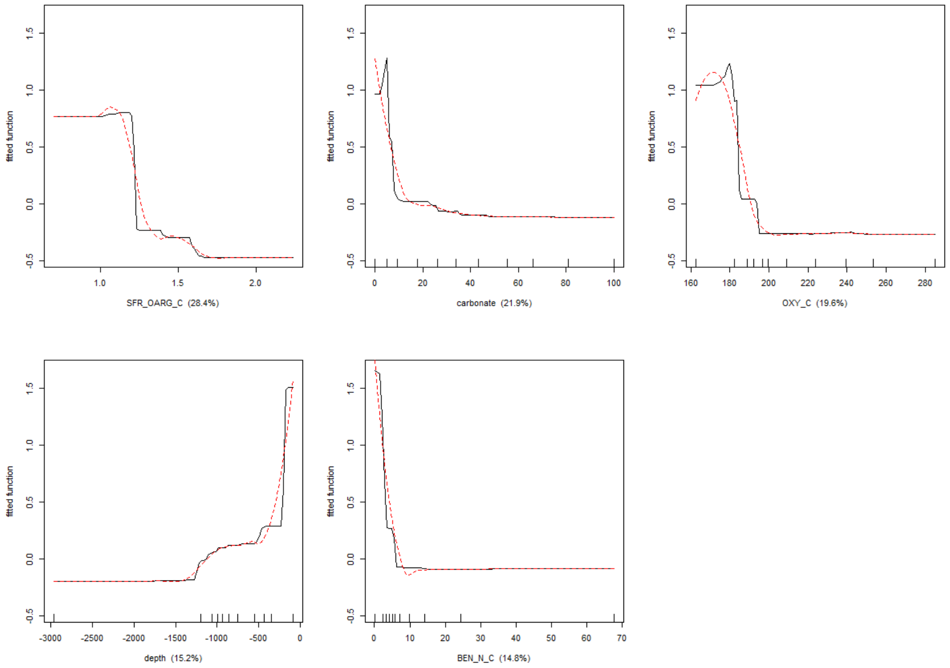
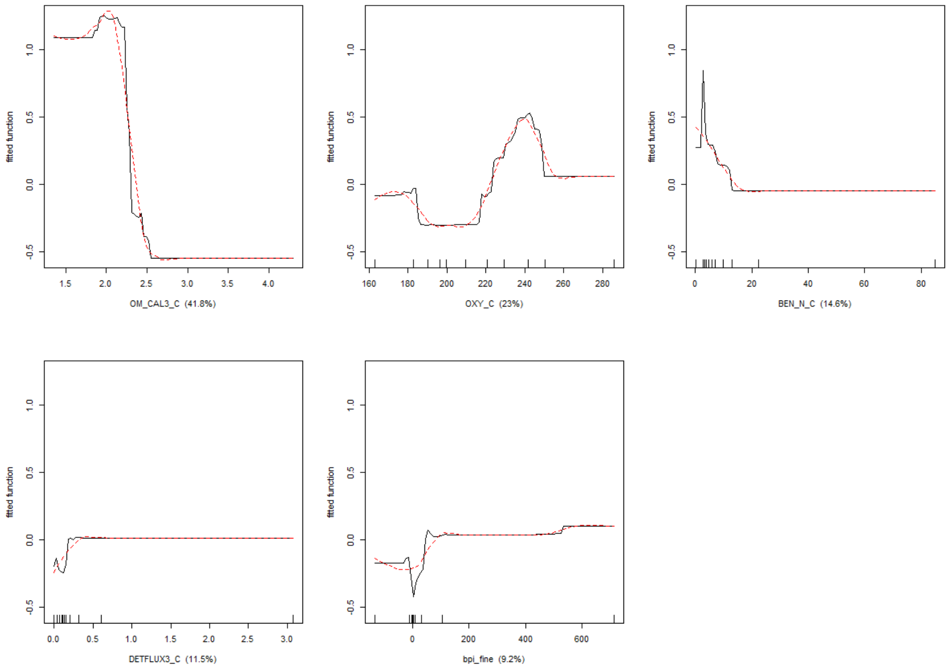
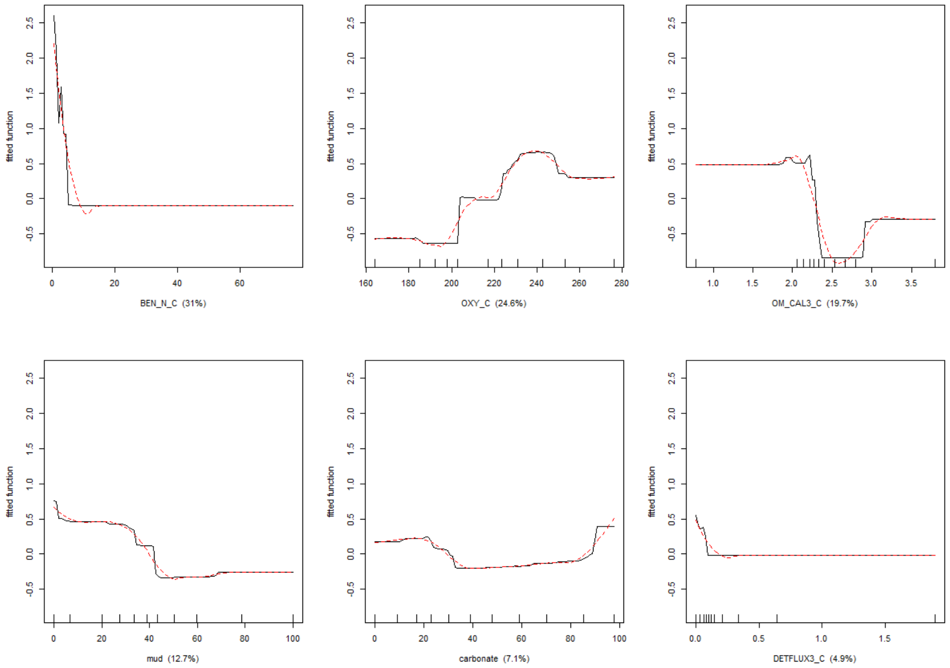


Figure C-2: Variable marginal effects in the Boosted Regression tree models for *Goniocorella dumosa* (top) and *Madrepora oculata* (bottom). See Table 2-2 for an explanation of the variables.



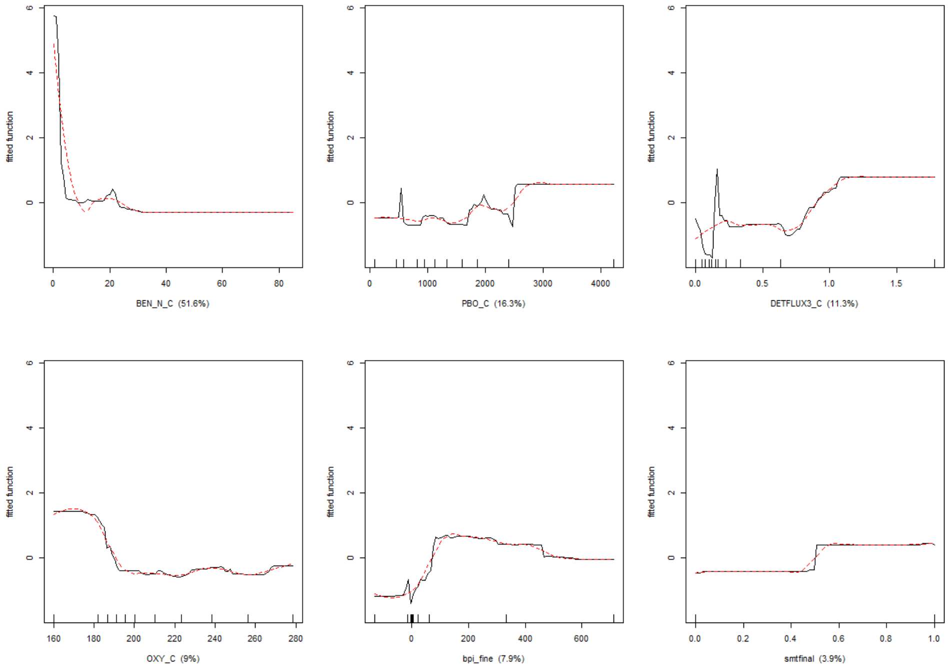
Paragorgia spp.



Primnoa spp.

Figure C-3: Variable marginal effects in the Boosted Regression tree models for *Paragorgia* spp. (top) and *Primnoa* spp. (bottom). See Table 2-2 for an explanation of the variables.

Corallium spp.



Keratoisis & *Lepidisis* spp.

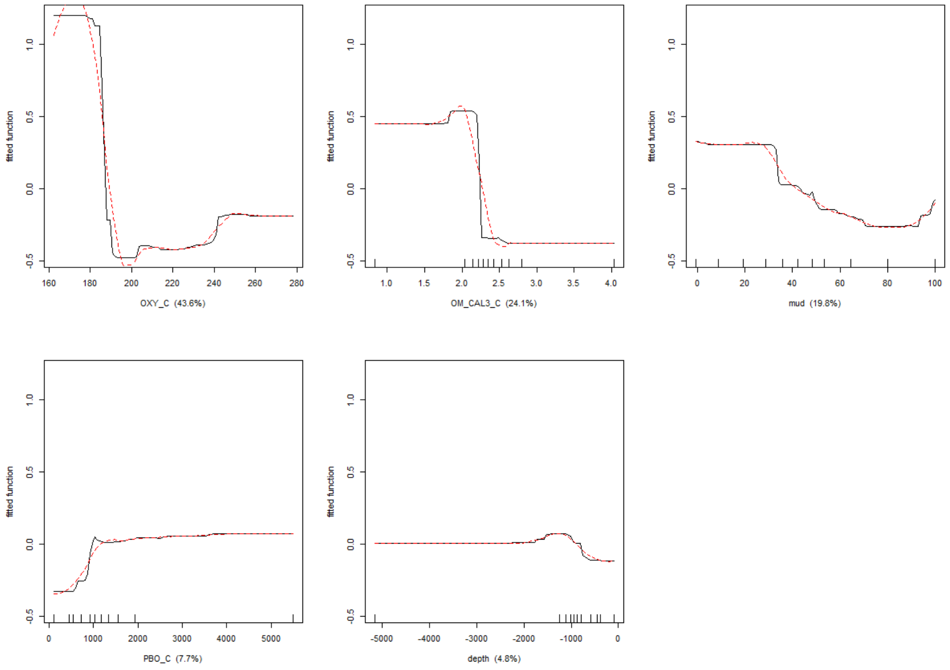
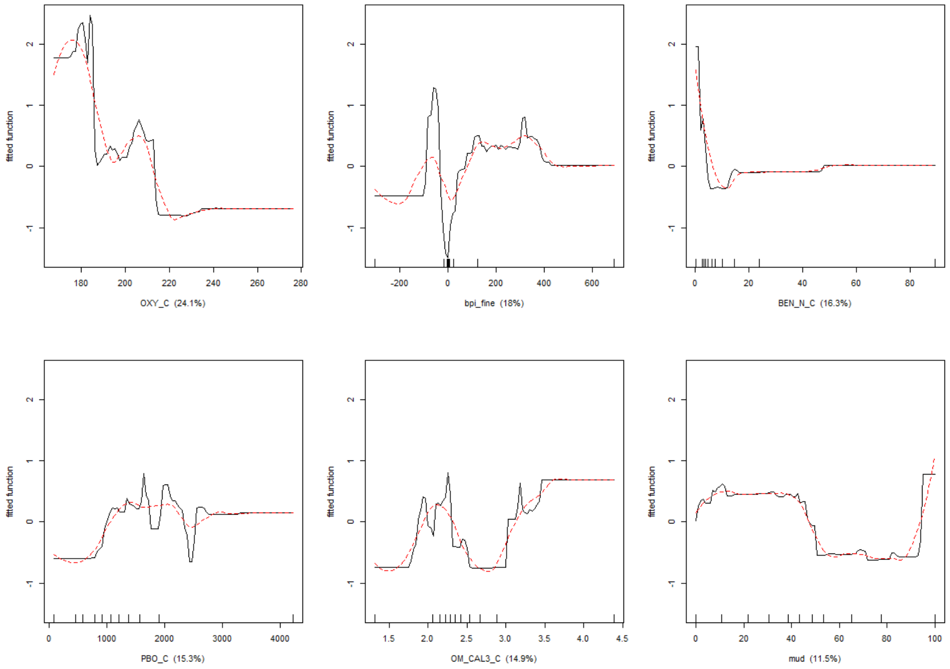


Figure C-4: Variable marginal effects in the Boosted Regression tree models for *Corallium* spp. (top) and *Keratoisis* & *Lepidisis* spp. (bottom). See Table 2-2 for an explanation of the variables.

Bathypathes spp.



Leiopathes spp.

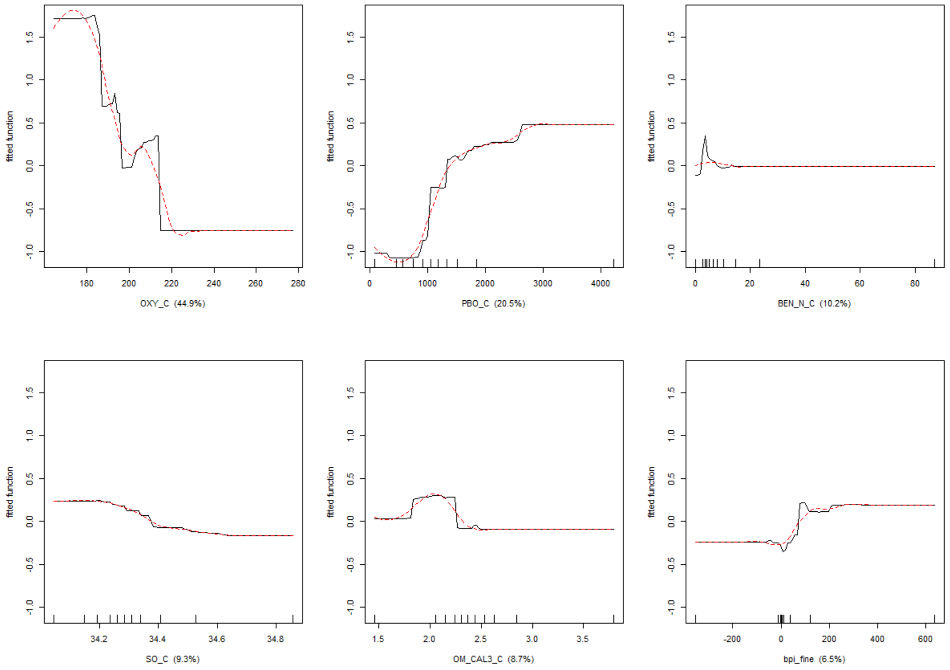
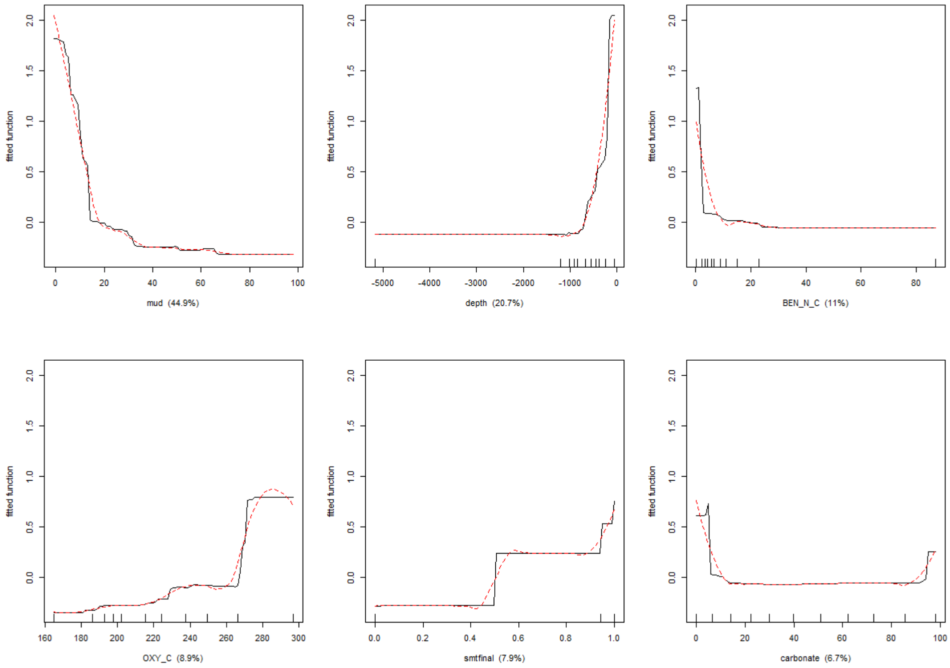
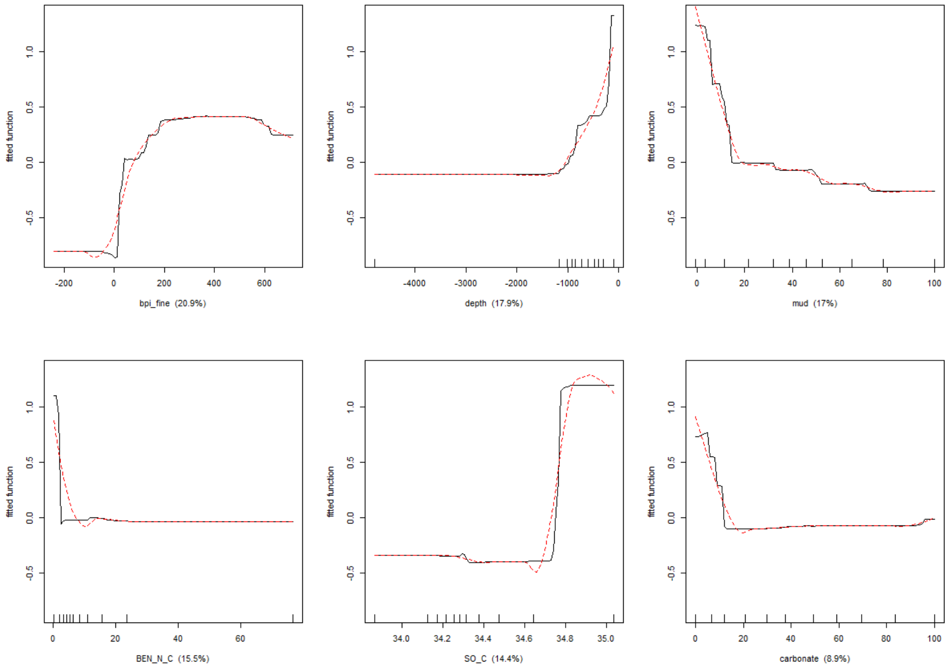


Figure C-5: Variable marginal effects in the Boosted Regression tree models for *Bathypathes* spp. (top) and *Leiopathes* spp. (bottom). See Table 2-2 for an explanation of the variables.



Errina spp.



Stylaster spp.

Figure C-6: Variable marginal effects in the Boosted Regression tree models for Bathypathes spp. (top) and Leiopathes spp. (bottom). See Table 2 2 for an explanation of the variables.