



*Supplement of*

## **An atlas of seabed biodiversity for Aotearoa New Zealand**

**Fabrice Stephenson et al.**

*Correspondence to:* Fabrice Stephenson ([fabrice.stephenson@waikato.ac.nz](mailto:fabrice.stephenson@waikato.ac.nz))

The copyright of individual parts of the supplement might differ from the article licence.

# Supplementary material 1

## Generating spatial predictions

The spatial distribution for each taxon was estimated using ensemble SDMs that were generated using the combined outputs from flexible machine learning Boosted Regression Tree (BRT) and Random Forest (RF) models. In subsequent sections we describe the biological data (from four biotic groups: demersal fish, reef fish, subtidal invertebrates and macroalgae – available from the authors upon reasonable request), the spatially explicit environmental data, and how these were combined to predict the taxa distributions used in the atlas of seabed biodiversity of Aotearoa New Zealand.

## Biological samples

### Demersal fish

Fish species records ( $n = 391,198$ ) (including information on research cruise identifier, gear type, date, minimum and maximum depth of trawl, and GPS location) from 1979 – 2016 were extracted from the research trawl database ‘TRAWL’ (Niwa, 2014, 2018). The data were groomed to only keep those records identified to species level, collected using bottom trawls and within the Aotearoa New Zealand Exclusive Economic Zone (EEZ) and Territorial Sea (TS). To minimise the effect of spatial bias in the occurrence data, species records were aggregated spatially to a 1 km grid resolution (Stephenson et al., 2020). Because of difficulties in correcting for differences in trawl methods, all catch records were converted into presence (Lundquist et al., 2020). To ensure distribution models were robust, only demersal fish species with  $\geq 50$  unique spatial locations were retained for analysis. The final dataset included presence/absence records of 235 demersal fish taxa at 28,599 unique sampling locations.

### Reef Fish

The relative abundance of reef fishes were obtained from 467 SCUBA dives made around the coast of Aotearoa New Zealand over an 18-year period from November 1986 to December 2004 (for detailed methodology see Smith et al. (2013)). The data were groomed for a previous study by Smith et al. (2013) and all records were provided to species level identification. Species records were aggregated (to presence/absence) spatially to a 250 m grid resolution and included observations of 160 species at 339 unique sampling locations. To ensure distribution models were robust, only reef fish species with  $\geq 35$  unique spatial locations were retained for analysis. The final dataset included presence/absence records of 51 reef fish taxa at 429 unique sampling locations.

### Subtidal Invertebrates

Subtidal invertebrate occurrence records ( $n = 127,330$ ) (including GPS location, species name, collection date, and sampling gear used) from 1896 – 2019 were extracted from TRAWL ( $n = 56,841$ ), NIWA invert ( $n = 59,144$ ), Te Papa ( $n = 2943$ ) and Auckland Museum ( $n = 8402$ ) databases. Only those records that had been classified to at least genus level and included information on sampling gear were extracted. Each record included information on the date, GPS location, survey and collection method. Across the four databases, 208 different methods were used to sample subtidal invertebrates, although many of these were name variants of commonly used sampling gears. To account for both the large number of gear types recorded and the differences in sampling parameters, gear types were grouped into catchability categories (Table S1). Catchability was assumed to be influenced by gear size, deployment area and selectivity (Stephenson et al., 2018b). Following categorisation of gear types, four gear classes were retained for species distribution modelling: SMG (small size, medium deployment area, general selectivity), SSG (small size, small deployment area, general selectivity), MMG (medium size, medium deployment

area, general selectivity) and LLG.LMG (Large size, medium and large deployment area, general selectivity). Gear class information was used to select representative absences for the modelled taxa, but samples from all gear types were combined for modelling. No taxa records collected using highly selective sampling methodologies were retained as these reflected opportunistic sampling which were not deemed comparable to the other methods used here.

Subtidal invertebrate records were spatially aggregated to a 1 km grid resolution. Genus level records were used because this provided a greater number of unique locations than when aggregated to species level (33,187 vs 28,263). To ensure distribution models were robust, only subtidal invertebrate genera with  $\geq 70$  occurrences were retained for analysis. The final dataset included records of 207 subtidal invertebrate taxa at 27,274 unique sampling locations.

**Table S1. Categories used to reflect catchability of sampling gear types. Table modified from (Stephenson et al. 2018b).**

Type	Category	Description	Example
Gear size	Small	< 1 m	Devonport dredge
	Medium	1-3 m	Benthic sled
	Large	> 3 m	Otter trawls
Deployment area	Small	< 1 m	Box corer
	Medium	10s to 100s of m	Beam trawls
	Large	> 1 km	Otter trawls
Selectivity	HS	Highly selective	Collected by hand
	G	General	Benthic sled

## Macroalgae

Macroalgae occurrence records based were extracted from herbaria housed at Te Papa Tongarewa - Museum of New Zealand, Auckland Museum Tāmaki Paenga Hira, and NIWA Taihoro Nukurangi. In addition, three observational datasets were included, representing 2,088 records. The first was based on citizen science observations of large brown algae, assembled as part of an Fisheries New Zealand (FNZ) funded project (ZBD201406). These citizen science contributions were verified via photographs with observation records submitted to NatureWatch. These observations extended the known/recorded distributions or filled in distributional gaps for a number of species (e.g., *Cystophora platylobium* - Kaikoura and Dunedin, *Cystophora retroflexa* - southern Hawkes Bay, *Cystophora scalaris* - East Cape, *Durvillaea poha* - Kaikoura and Stewart Island, *Hormosira banksii* - Wairarapa and north Otago, *Macrocystis pyrifera* - SE Otago). The second was extracted from dive logs contributed by Clinton Duffy (Department of Conservation, Auckland) of large brown seaweed observed around New Zealand between 1979 and 2007. The third was data collected by Shears & Babcock (2007) during their work on shallow subtidal reef communities.

The material in natural history collections and herbaria only provide presence data, establishing that the species was present at that locality when collected, and the interpretation of species absences is complex, i.e., the species may not have been at the locality, or was not collected, or not detected. Species records were aggregated spatially to a 250 m grid resolution and included observations of 349 species at 3,320 unique sampling locations. To ensure distribution models were robust, only macroalgae species with  $\geq 50$  unique spatial locations were retained for analysis. The final dataset included presence records of 86 macroalgae taxa at 2603 unique sampling locations.

## Environmental conditions

A comprehensive number of spatial environmental variables data for the Aotearoa New Zealand EEZ/TS were available from Stephenson et al. (2022a) (Table S2). These variables were selected because they had previously been identified as being important predictors of the distributions of the taxa used here (Stephenson et al., 2022).

Several of the environmental variables are scheduled for continuous review and updating, to incorporate new data and ensure the use of the most up-to-date analysis routines for satellite remote sensed data. When a number of new datasets are available, this will trigger a refitting of the SDMs represented in this atlas, however, it is not anticipated that small changes in environmental data will cause dramatic shifts in predicted species distribution.

**Table S2. Description of spatial environmental predictor variables ( $n=20$ , in alphabetical order) collated for species distribution models (table modified from Stephenson et al. (2022)).**

Abbreviation	Full name	Description	Units	Source
<i>Bathy</i>	Bathymetry	Depth at the seafloor was interpolated from contours generated from various sources, including multi-beam and single-beam echo sounders, satellite gravimetric inversion, and others (Mitchell et al., 2012).	m	Mitchell et al. (2012)
<i>Beddist</i>	Benthic sediment disturbance	One-year mean value of friction velocity derived from (1) hourly estimates of surface wave statistics (significant wave height, peak wave period) from outputs of the NZWAVE_NZLAM wave forecast, at 8-km resolution, (2) median grain size (d50), at 250 m resolution, (3) water depth, at 25-m resolution. Benthic sediment disturbance from wave action was assumed to be zero where depth $\geq$ 200 m.	ms <sup>-1</sup>	Swart (1974); updated in 2019
<i>BotNi</i>	Bottom nitrate	Annual average water nitrate concentration at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from Dunn et al. (2002). The oceanographic data used to generate these climatological maps were computed by objective analysis of all scientifically quality-controlled historical data from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atlas of Regional Seas database (Cars2009, 2009).	umol l <sup>-1</sup>	NIWA, unpublished
<i>BotOxy</i>	Dissolved oxygen at depth	Annual average water oxygen concentration at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from (Ridgway et al., 2002). The oceanographic data used to generate these climatological maps were computed by objective analysis of all scientifically quality-controlled historical data from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atlas of Regional Seas database (Cars2009, 2009).	ml l <sup>-1</sup>	NIWA, unpublished

<i>BotPhos</i>	Bottom phosphate	Annual average water phosphate concentration at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from Dunn et al. (2002). The oceanographic data used to generate these climatological maps were computed by objective analysis of all scientifically quality-controlled historical data from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atlas of Regional Seas database (Cars2009, 2009)	umol l <sup>-1</sup>	NIWA, unpublished
<i>BotSal</i>	Salinity at depth	Annual average water salinity concentration at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from (Ridgway et al., 2002). The oceanographic data used to generate these climatological maps were computed by objective analysis of all scientifically quality-controlled historical data from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atlas of Regional Seas database (Cars2009, 2009)	psu	NIWA, unpublished
<i>BotSil</i>	Bottom silicate	Annual average water silicate concentration at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from (Ridgway et al., 2002). The oceanographic data used to generate these climatological maps were computed by objective analysis of all scientifically quality-controlled historical data from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atlas of Regional Seas database (Cars2009, 2009)	umol l <sup>-1</sup>	NIWA, unpublished
<i>BotTemp</i>	Temperature at depth	Annual average water temperature at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from (Ridgway et al., 2002). The oceanographic data used to generate these climatological maps were computed by objective analysis of all scientifically quality-controlled historical data from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atlas of Regional Seas database (Cars2009, 2009).	°C km <sup>-1</sup>	NIWA, unpublished
<i>BPI_broad</i>	BPI_broad	Terrain metrics were calculated using an inner annulus of 12 km and a radius of 62 km using the NIWA bathymetry layer in the Benthic Terrain Modeler in ArcGIS 10.3.1.1 (Wright et al., 2012). Bathymetric Position Index (BPI) is a measure of where a referenced location is relative to the locations surrounding it.	m	NIWA, unpublished
<i>BPI_fine</i>	BPI_fine	Terrain metrics were calculated using an inner annulus of 2 km and a radius of 12 km using the NIWA bathymetry layer in the Benthic Terrain Modeler in ArcGIS 10.3.1.1 (Wright et al., 2012). Bathymetric Position Index (BPI) is a measure of where a referenced location is relative to the locations surrounding it.	m	NIWA, unpublished

<i>Chl a</i>	Chlorophyll-a concentration	A proxy for the biomass of phytoplankton present in the surface ocean (to ~30 m). Blended from a coastal Chl-a estimate (quasi-analytic algorithm (QAA), local $aph^*(555)$ ) and the default open-ocean chl-a value from MODIS-Aqua (v2018.0).	$mg\ m^{-3}$	NIWA unpublished; Based on processing described in (Pinkerton, 2016) and updated in (Pinkerton et al., 2020).
<i>Chl-a.Grad</i>	Chlorophyll-a concentration spatial gradient	Smoothed magnitude of the spatial gradient of annual mean chlorophyll-a concentration. Derived from Chl-a above.	$Mg\ m^{-3}\ km^{-1}$	NIWA unpublished, updated in 2020; Based on processing described in (Pinkerton et al., 2018)
<i>DET</i>	Detrital absorption	Total detrital absorption coefficient at 443 nm, including due to coloured dissolved organic matter (CDOM) and particulate detrital absorption. Estimated using quasi-analytic algorithm (QAA) applied to MODIS-Aqua data, blended with <i>adg_443_giop</i> ocean product (Werdell, 2019).	$m^{-1}$	NIWA unpublished, updated in 2020; Based on processing described in (Pinkerton et al., 2018). Processing for <i>adg_443_giop</i> ocean product described in (Werdell, 2019).
<i>Ebed</i>	Seabed incident irradiance	Broadband (400–700 nm) incident irradiance ( $E\ m^{-2}\ d^{-1}$ ) at the seabed, averaged over a whole year. Estimated by combining incident irradiance at the sea surface ((Frouin et al., 2012); this table), diffuse downwelling irradiance attenuation ( $K_{PAR}$ ; this table) and bathymetric depth at monthly resolution. Derived from blended coastal (QAA) and open-ocean attenuation products.	$E\ m^{-2}\ d^{-1}$	NIWA unpublished, updated in 2020, based on processing described in Pinkerton et al. (2018)
<i>POCFlux</i>	Downward vertical flux of particulate organic matter at the seabed	Net primary production in the surface mixed layer estimated as the VGPM model ((Behrenfeld and Falkowski, 1997); this table). Export fraction and flux attenuation factor with depth estimated by refitting sediment trap and thorium-based measurements to environmental data (VGPM, SST) as (Lutz et al., 2002), Pinkerton et al. (2016) and using data from (Cael et al., 2018).	$mgC\ m^{-2}\ d^{-1}$	NIWA unpublished, updated in 2020. Based on processing described in Pinkerton et al. (2016) with new data from Cael et al. (2018).
<i>PB555nm</i>	Particulate backscatter at 555 nm (previously used to generate 'turbidity')	Optical particulate backscatter at 555 nm estimated using blended coastal and ocean products. Coastal: QAA v5 product <i>bbp555</i> from MODIS-Aqua data. Ocean: <i>bbp_555_giop</i> ocean product (Werdell, 2019). Result calculated as long-term (2002–2017) average.	$m^{-1}$	NIWA unpublished, updated in 2020; Based on processing described in Pinkerton et al. (2018). Processing for <i>bbp_555_giop</i> ocean product described in Werdell (2019).

<i>SeasTDiff</i>	Annual amplitude of sea floor temperature	Smoothed difference in seafloor temperature between the three warmest and coldest months. Providing a measure of temperature amplitude through the year.	°C km <sup>-1</sup>	NIWA, unpublished data, updated in 2018
<i>Sed.class</i>	Sediment classification	Classification of Mud, Sand and Gravel layers (this table) using the well-established (Folk et al., 1970) classification. Subtidal rocky reefs (this table) were incorporated. This classification provides a broad measure of hardness Mud – Rock.	NA; Mud; Muddy gravel; Muddy sandy gravel; sand; Gravelly mud; Gravelly sandy mud; Gravelly sand; Gravel; Rock	NIWA unpublished, updated in 2020
<i>Slope</i>	Slope	Bathymetric slope was calculated from water depth and is the degree change from one depth value to the next.	°	NIWA, unpublished, updated in 2019
<i>SSTGrad</i>	Sea surface temperature gradient	Smoothed magnitude of the spatial gradient of annual mean SST. This indicates locations in which frontal mixing of different water bodies is occurring (Leathwick et al., 2006). Derived from SST described above at two resolutions and merged.	°C km <sup>-1</sup>	NIWA unpublished, updated in 2020
<i>TC</i>	Tidal Current speed	Maximum depth-averaged (NZ bathymetry) flows from tidal currents calculated from a tidal model for New Zealand waters (Walters et al., 2001). Tidal constituents (magnitude A and phase phi, represented as real and imaginary parts $X + iY = A \cdot \exp(i \cdot \phi)$ ) for sea surface height and currents (8 components) were taken from the EEZ tidal model, on an unstructured mesh at variable spatial resolution. The complex components were bilinearly interpolated to the output grid.	ms <sup>-1</sup>	Walters et al., 2001; NIWA unpublished, updated in 2020

### Species Distribution Models (SDMs)

Ensemble predictions from Boosted Regression Tree (BRT) and Random Forest (RF) species distribution models (Ensemble SDMs) were produced for each taxon. 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 (Stephenson et al., 2021).

### Random Forest models

Random Forest models (Breiman, 2001) fit an ensemble of regression (abundance data) or classification tree (presence/absence data) models describing the relationship between the distribution of an individual species and some set of environmental variables (Ellis et al., 2012). Following environmental predictor selection using an initial RF model, a second RF model was tuned using additional R routines based on methodologies previously applied to subtidal invertebrate data (Rowden et al., 2017; Georgian et al., 2019; Stephenson et al., 2021) and demersal fish in the New Zealand region (Stephenson et al., 2018a).

### Boosted Regression Tree models

BRT modelling combines many individual regression trees (models that relate a response to their predictors by recursive binary splits) and boosting (an adaptive method for combining many simple models to give improved predictive performance) to form a single ensemble model (Elith et al., 2008). Detailed descriptions of the BRT method are available in Ridgeway (2007) and Elith et al. (2008). All statistical analyses were undertaken in R (R Core Team, 2020) using the ‘Dismo’ package (Hijmans et

al., 2017). BRT models were fitted with a Bernoulli error distribution, a tree complexity of 2 – 3, a learning rate between 0.01 – 0.0001 (with parameters selected so as to fit between 1000 and 3000 trees for each species' model), a bag fraction of 0.6 and random 10-fold cross evaluation following recommendations from Elith et al. (2008) and Leathwick et al. (2006).

The BRT method has been widely used in ecological applications and has performed well in previous studies of invertebrate and fish distributions in Aotearoa New Zealand (Leathwick et al., 2006; Compton et al., 2013; Anderson et al., 2016; Stephenson et al., 2021).

### Species Distribution Model tuning

To estimate taxonomic distributions, BRT and RF models require locations of both presences (occurrence records) and absences. For biotic groups demersal fish and reef fish, the consistent methods used to collect occurrence data means that where a taxon was not sampled, we assume that taxon was absent. For biotic group subtidal invertebrates, the same assumption was made but split by gear type (to account for differences in sampling efficiency). For biotic group macroalgae, given the differences in sampling methods used to collect occurrence data, we used 'target-group background data' (Phillips et al., 2009) as absences (referred to here as relative absence), i.e., a location where a different taxon to that being modelled was recorded (Stephenson et al., 2020). In practice different labelling of 'absences' does not affect the modelling approach but illustrates differences in the certainty of the absences (and therefore the outputs). Absence (and relative absences for macroalgae) were generated for each taxon from occurrences within taxonomic groups (i.e., demersal fish absences were generated from demersal fish occurrence records). The location of absences and relative absences was required to be at least 1 km from presence data and the number of absence and relative absence data was set to be equal to the number of presences (following best practice outlined in Aiello-Lammens et al. (2015) and Barbet-Massin et al. (2012)).

In most cases, the inclusion of many variables (e.g., all 20 + variables) is avoided because they generally only provide minimal improvement in predictive accuracy, and complicate interpretation of model outcomes (Leathwick et al., 2006). Several environmental variables showed some co-linearity within records for taxa groups; however, all levels of co-linearity were considered acceptable (Pearson correlation < 0.9) for tree-based machine learning methods (Elith et al., 2010; Dormann et al., 2013). To produce parsimonious models, an automated environmental variable selection was performed. In the first instance a RF model was fitted to the presence / target-group background data using the extended Forest package in R (Liaw and Wiener, 2002). This method accounts for any co-linearity in environmental predictor variables when determining the relative importance of each predictor variable in the model through the implementation of a conditional approach to variable importance calculation (Ellis et al., 2012). Only environmental variables with a relative influence > 5% were retained (Müller et al., 2013; Jouffray et al., 2019). This allowed environmental predictors that may have important localised importance, but with low overall importance, to be retained whilst removing any very low, or negatively contributing environmental variables (R Pitcher, pers. comm.). For each taxon, the 'final' environmental variables selected through this approach were also used in the BRT models.

BRT and RF models were bootstrapped 100 times for each taxon modelled. That is, we selected random 'training' samples with a sample size equal to the number of presence records, and an equal number of randomly selected target group background data records, both with replacement. The bootstrapping process was repeated 100 times, and at each iteration, predictions were made to the 'evaluation' data (i.e., the remaining presence data not randomly selected and an equal number of randomly selected target group background data records).

At each BRT and RF model iteration, geographic predictions were made using environmental predictor variables to a 1 km<sup>2</sup> grid. For each taxon, habitat suitability index (HSI) and a spatially explicit measure of uncertainty (measured as the standard deviation of the mean (SD)) were calculated for each grid cell using the 100 bootstrapped layers. Spatial predictions of HSI represent the predicted relative suitability of the environmental envelope for each taxon ranging from 0 to 1, with higher values indicating a more suitable environment (Georgian et al., 2019). While such outputs are frequently presented as "probability of



135 occurrence” we use the more appropriate term “HSI” because of the lack of true absence data and unknown efficiency of the  
 sampling gear (Stephenson et al., 2021). To aid interpretation of HSI values we follow the subjectively defined categorization  
 by Georgian et al. (2019):  $< 0.4$  = low suitability,  $0.4 - 0.8$  = moderate suitability, and  $> 0.8$  = high suitability. To avoid  
 predictions into unsampled space (e.g., into deep areas with few biological samples), geographic predictions were clipped to  
 140 of reef habitat for reef fish.

### Ensemble models

Ensemble models were produced for each taxon by taking weighted averages of the predictions from each model type, using  
 methods described in Stephenson et al. (2021) and Anderson et al. (2020). This adapted procedure derives a two-part weighting  
 145 for each component of the ensemble model, taking equal contributions from the overall model performance (AUC) of each  
 mode type (BRT / RF) and the uncertainty measure (SD) in each cell from spatial predictions from each model type, as follows:

$$W_{1BRT} = \frac{MPS_{BRT}}{MPS_{BRT} + MPS_{RF}} \text{ and } W_{1RF} = \frac{MPS_{RF}}{MPS_{BRT} + MPS_{RF}}$$

$$W_{2BRT} = 1 - \frac{SD_{BRT}}{SD_{BRT} + SD_{RF}} \text{ and } W_{2RF} = 1 - \frac{SD_{RF}}{SD_{BRT} + SD_{RF}}$$

150

$$W_{BRT} = \frac{W_{1BRT} + W_{2BRT}}{2} \text{ and } W_{RF} = \frac{W_{1RF} + W_{2RF}}{2}$$

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

$$155 \quad SD_{ENS} = SD_{BRT} * W_{BRT} + SD_{RF} * W_{RF}$$

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

### 160 Species distribution model performance

An important component of being able to use the predicted distributions adequately for management or future scientific  
 investigations is to estimate how well the models perform. That is, if a model is assessed as performing very poorly, it may be  
 best to use the associated prediction with caution or not at all. Historically, this model assessment was derived from “statistical  
 model fit metrics” (i.e., assessing how well the model predicts to the available biological data). However, these measures aren’t  
 165 perfect and can sometimes result in over-confidence in the model results (Bowden et al., 2021). Here we present methods used  
 to assess model accuracy using statistical model fit metrics and additionally a separate expert assessment. Combined, these  
 assessments provide complementary information and will ensure that weaknesses or errors in the predicted taxon’ layers are  
 highlighted.

Ensemble model performance was assessed using AUC (area under the Receiver Operating Characteristic curve) and True  
 Skill Statistic (TSS, which takes into account Specificity and Sensitivity to provide an index ranging from -1 to +1, where +1  
 170 equals perfect agreement and -1 = no better than random, (Allouche et al., 2006)) using all the taxon’s presence data and an  
 equal number of randomly selected target-group background data (Table S3). To ensure that the random selection of target-  
 group background data did not provide misleading model performance metrics, this procedure was iterated 50 times and mean  
 AUC and TSS score calculated for the ensemble model (Barbet-Massin et al. 2012).

175 **Table S3. Description and evaluation score of taxon statistical model**

<b>Evaluation score</b>	<b>Description</b>	<b>Details</b>
1 – Excellent	Model metrics on evaluation data are high and large number of records	TSS on evaluation data > 0.75 for models and > than 200 records
2 – Good	Model metrics on evaluation data are high	TSS on evaluation data > 0.75 for models
3 – Moderate	Model metrics are moderate	TSS on evaluation data > 0.3 for models
4 – Poor	Model metrics are low	TSS on evaluation data < 0.3 for models

Predicted distributions were also evaluated by taxonomists and ecologists with expertise in the predicted taxon’s distributions. Experts’ evaluation was elicited during several workshops in October 2021. Each taxon’s distribution was evaluated following scoring of three sequentially assessed evaluation metrics: (1) assessment of experts’ knowledge of taxon’s distribution (i.e., does the expert feel confident judging the validity of the predicted distribution); (2) assessment in the congruence between the distribution of available taxon records and expert’s view of the distribution; and (3) assessment in the congruence between predicted taxon distribution and expert’s view of taxon distribution (Tables S4, S5 and S6). The first evaluation metric (Table S4) aimed to reflect the level of confidence that the expert had in assessing the distribution of each taxon, i.e., this can be viewed as a measure of uncertainty of the expert evaluation. The second evaluation metric (Table S5) was used to assess whether taxon records were distributed in a manner that was likely to reflect the taxon’s distribution (as assessed by the expert). This evaluation metric was particularly useful for highlighting possible spatial bias in the distribution of the taxon records (i.e., sampling and subsequent records were collected in the absence of a predefined sampling scheme often influenced by accessibility, higher potential of observations, or previous knowledge of the study area (Glad et al., 2019)) as this can result in model predictions that do not accurately reflect taxa’ environmental preferences. Finally, the third evaluation metric (Table S6) was used to assess whether the overall predicted taxon distribution reflected the expert’s view of the taxon’s distribution (assumed to be an accurate representation based on their scoring of their knowledge of the taxa’ distribution). Only the final expert model evaluation is provided in the online metadata. That is, the expert evaluation score, which in combination with the TSS statistical model evaluation, are assumed to be the most useful metrics for assessing the robustness of the predicted distributions. However, all other statistical and expert evaluations scores are provided for each taxon in Supplementary materials 2 (Stephenson et al., 2022b; <https://doi.org/10.5281/zenodo.6781435>). In most cases, statistical and expert validations were congruent. For example, the predicted distributions of kahawai (*Arripis trutta* – demersal fish), the erect branching deep-water coral genus *Solenosmilia* sp. (subtidal invertebrate); the wrasse *Pseudolabrus miles* (reef fish); and the New Zealand bull kelp (*Durvillaea antarctica*, macroalgae) were all considered to be “accurate” / “very accurate” by experts and had “excellent” statistical model validation scores. In contrast, there were some taxa for which the expert evaluation scores were much lower (“inaccurate” / “somewhat inaccurate”) than the statistical evaluation score (“good”). For example, the pacific salmon (*Oncorhynchus tshawytscha*, demersal fish), the bryozoan Genus *Figularia* (subtidal invertebrates), the wrasse *Coris sandeyeri* (reef fish), the red algae *Champia novae-zelandiae* (macroalage). Care must be taken when using predictions where there is discrepancy between statistical and expert evaluation (with expert evaluation assumed to be more accurate).

205 **Table S4. Description and evaluation score of expert knowledge of the taxon’s distribution**

<b>Evaluation score</b>	<b>Description</b>
1 – Very high	Expert confidently knows the distribution of the taxon (including relatively fine-scale patterns)
2 – High	Expert confidently knows the broadscale distribution of the taxon
3 – Moderate	Expert has some knowledge of the likely distribution of the taxon with some uncertainty
4 – Low	Expert has little knowledge of likely distribution of the taxon and with large uncertainty

**Table S5. Description and evaluation score of congruence between the distribution of taxon records and the expert’s view of the taxon’s distribution**

Evaluation score	Description
1 – Very accurate	Distribution of taxon records reflect expert’s view of the taxon’s distribution (> 80% overlap).
2 – Accurate	Distribution of taxon records reflect expert’s view of the taxon’s distribution, but some areas do not (> 60% overlap).
3 – Somewhat accurate	Distribution of taxon records somewhat reflect expert’s view of the taxon’s distribution but there are considerable inconsistencies (i.e. large areas that do not reflect expert’s views; > 40% agreement) – moderate spatial bias in records.
4 – Inaccurate	Distribution of taxon records do not reflect expert’s view of the taxon’s distribution (< 40% agreement) – high spatial bias in records.

210 **Table S6. Description and evaluation score of congruence between predicted species distribution and expert view of species distribution**

Evaluation score	Description
1 – Very accurate	Predicted distribution reflects expert view of taxon’s distribution (> 80% overlap)
2 – Accurate	Predicted distribution reflects expert view of taxon’s distribution, but some areas may not be correct (> 60% overlap)
3 – Somewhat accurate	Predicted distribution somewhat reflects expert view of the taxon’s distribution but there are considerable inconsistencies (i.e., regions of disagreement; > 40% agreement)
4 – Largely inaccurate	Predicted distribution contains large inconsistencies with the expert’s view of the taxon’s distribution (i.e., large regions of disagreement; > 20% agreement)
5 – Inaccurate	Predicted distribution does not match the expert’s view of the taxon’s distribution (i.e., < 20% agreement)

### Metadata for species distribution layers

215 Metadata for each taxon’s distribution layer comprised 17 categories. Metadata categories, description of categories and an example for the demersal fish *Chrysophrys auratus* (Australasian snapper) is provided in Table S7. Metadata for all taxa are available on the Aotearoa New Zealand Department of Conservation Marine Data Portal (<https://doc-marine-data-deptconservation.hub.arcgis.com>).

**Table S7. Metadata categories, description of categories and an example for the demersal fish *Chrysophrys auratus* (Australasian snapper)**

Category	Description	Example
File name	The name of the raster file representing species' distributions provided as Geotiff (.tif)	Chrysophrys.auratus_ENS.tif
Link (file path)	File path	Q:\Data\Marine\Key_Ecological_Areas\KEA_GIS_Master_database_2020\Species_layers\Fish\Demersal_Fish_Models_2020\Demersal_Fish_modelled_occurrence_239spp\C_hrysophrys.auratus_ENS.tif
Scientific name	Scientific name (up-to-date taxonomy as of 2021 based on WoRMs database)	Chrysophrys auratus
Title	Title of the layer	Predicted geographical distribution of <i>Chrysophrys auratus</i> (demersal fish)
Thumbnail		
Tags	Other searchable names for the taxa	Snapper , Chrysophrys [Pagrus]auratus , Karati, Taamure
Summary	Summary of the layer	Predicted geographic distribution of <i>Chrysophrys auratus</i> (demersal fish) -

Description	Detailed description of the layer including important metrics and scores for model evaluation and expert evaluations.	<p>measured as habitat suitability index (HSI: 0-1).</p> <p>Predicted geographic distribution of <i>Chrysophrys auratus</i> (demersal fish) using ensemble Species Distribution Modelling (Bootstrapped Boosted Regression Tree and Random Forest models) described in Lundquist et al., 2020. Spatial predictions generated for all geographic areas within the EEZ to depths of 2000 m (areas considered to have adequate sample coverage). Associated spatially explicit uncertainty predictions are available for this taxon (see layer ‘Uncertainty for Predicted Geographical Distribution of <i>Chrysophrys auratus</i> (demersal fish)’).</p> <p>Number of taxa records: 2932</p> <p>Statistical model performance: Good (TSS = 0.88)</p> <p>Expert evaluation of predicted geographical distribution: 1, very accurate</p> <p>Spatial resolution: 1 km</p>
Credits	Credits and reference	<p>This database was prepared to support the Key Ecological Areas (KEA) geospatial database by Fabrice Stephenson and Tom Brough see Stephenson &amp; Brough et al. 2022.</p> <p>Data are summarised in Lundquist, C.; Stephenson, F.; McCartain, L.; Watson, S.; Brough, T.; Nelson, W.; Neill, K.; Anderson, T.; Anderson, O.; Bulmer, R.; Gee, E.; Pinkerton, M.; Rowden, A.; Thompson, D. 2020. Evaluating Key Ecological Areas datasets for the New Zealand Marine Environment. NIWA Client report No.2020109HN, prepared for Department of Conservation (project DOC19206). 120 p.</p>
Use limitations	Specific conclusions on the use of the layer based on expert validation model evaluation scores	No specific limitations (expert score 1-3)
Date - Created	Date the layer was generated	2020
Update frequency	Whether (and when) the layer is planned to be updated	No updates planned

General Constraints	General disclaimers and constraints	Geographical predictions represent habitat suitability index and do not necessarily reflect spatial patterns of abundance. In addition, the predicted distribution may have inaccuracies and must not be interpreted as the true distribution of <i>Chrysophrys auratus</i> (demersal fish) but rather as an estimate reflecting the information available at the time of the analysis.
Validity period		Taxonomic nomenclature valid as of 2021 and reviewed by the demersal fish expert working group (further information on the expert working is available upon request).
Licence	Access licence type for the layer	CC-BY
Lineage	Original provenance of the layer	Species occurrence records and environmental co-variates used in the species distribution models were compiled from various open sources and are described in Stephenson et al. (2022).
		Further information on the expert evaluation process and results is available upon request.
		The predicted geographic distribution of demersal fish presented here supersedes those presented in Leathwick et al. (2006) and Stephenson et al. (2018).
		Leathwick, J., Elith, J., Francis, M., Hastie, T., and Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. <i>Marine Ecology Progress Series</i> 321, 267-281
		Stephenson, F., Rowden, A., Anderson, T., Hewitt, J., Costello, M., Pinkerton, M., Morrison, M., Clark, M., Wadhwa, S., Mouton, T., and Lundquist, C. (2018). Mapping Key Ecological Areas in the New Zealand Marine Environment: Data collation. NIWA Client Report 2018332HN, prepared for the Department of Conservation (DOC).

Contact

Contact details for further enquiries

Marine Ecosystems Team, Department of  
Conservation Te Papa Atawhai  
[marine@doc.govt.nz](mailto:marine@doc.govt.nz)

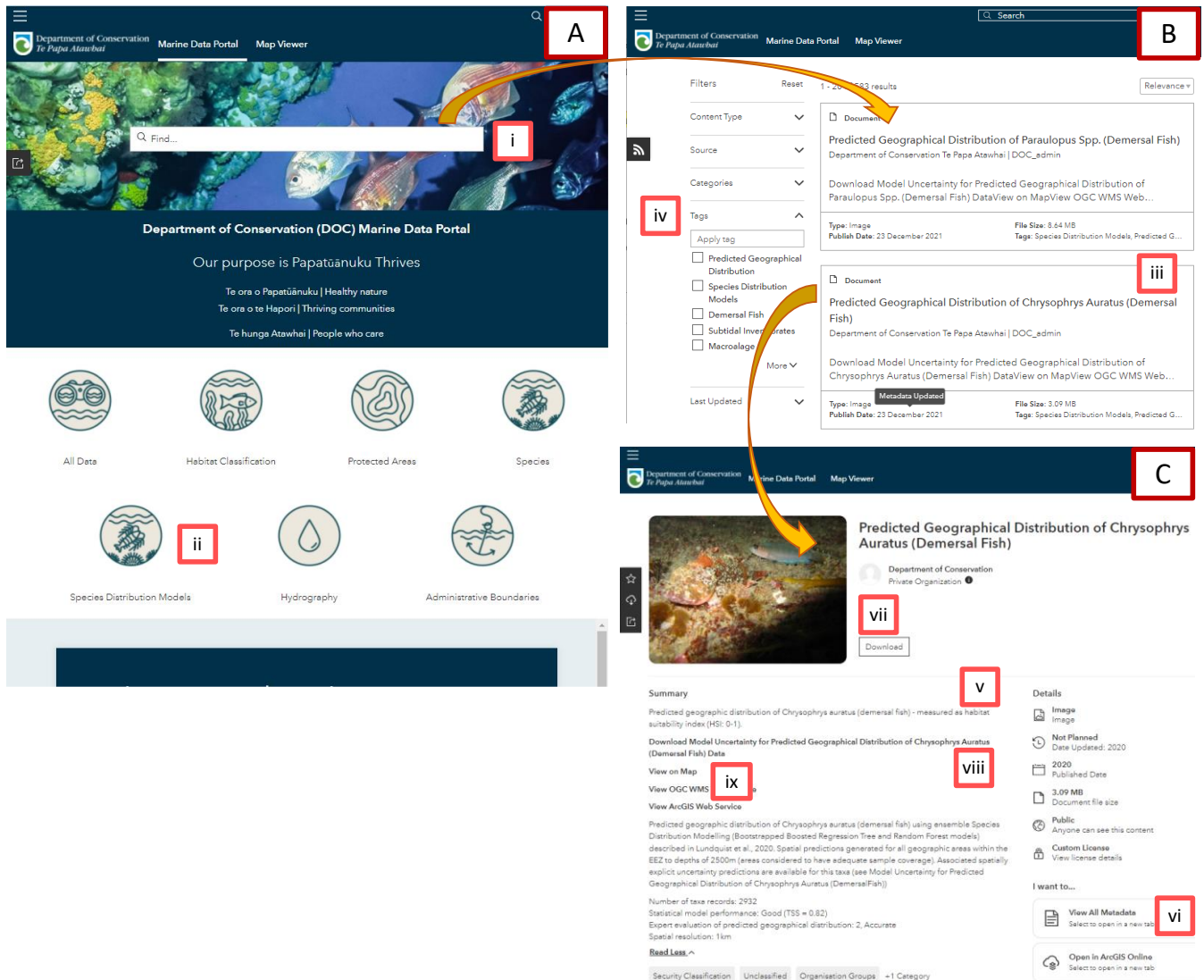
---

220

# Supplementary material 2

## The online database

225 Taxa distributions and associated uncertainty estimates can be searched for, and viewed, by inputting the taxa name, or for  
common taxa, using their vernacular name (Figure S1, i). For example, Australasian snapper (*Chrysophrys auratus*) can be  
searched using “Chrysophrys”, “auratus”, “Chrysophrys auratus”, “snapper”, “Karati” and “Taamure”. Alternatively, all taxa  
distributions can be viewed in the online database (Figure 3, A) by clicking on the “Species Distribution Models” clickable  
box (Figure S1, ii). The “Tag” tab (Figure S1, iv) can be used to further refine the available layers (e.g., all demersal fish layers  
can be selected). Searches by species name can be performed in the “Tag” tab or the “Search bar”. For an easy overview of  
230 the available taxa layers (i.e., taxonomic and vernacular taxa names) and statistical and expert evaluation scores see  
Supplementary materials 2.



235 **Figure S1.** Illustration of database features. (A) The Department of Conservation Marine Data Portal front page; (B) the online atlas and search functionality; (C) Metadata and links to the prediction distribution and associated uncertainty for Australasian snapper (*Chrysophrys auratus*, a demersal fish species).

## 3.2 Metadata

All the relevant information (metadata) is available by clicking on the link for individual taxa (Figure S1, iii). The summary (Figure S1, C) provides key information on the data layer, including an overview of the data and methods used for generating

240 the layer and the statistical and expert evaluation scores (Figure S1, v). Other metadata are available by clicking on the “View  
All Metadata” clickable box (Figure S1, vi), where information on model development/validation (see the *abstract* field) is  
combined with standard metadata reporting. The layer can be downloaded as a raster file (saved as a GeoTiff) for use in GIS  
or further analysis by clicking the “Download” button (Figure S1, vii), and the spatially explicit uncertainty layer can be  
downloaded by clicking the “Download Model Uncertainty for Predicted Geographical Distribution of \*taxon name\* Data”  
245 link in the summary (Figure S1, viii).

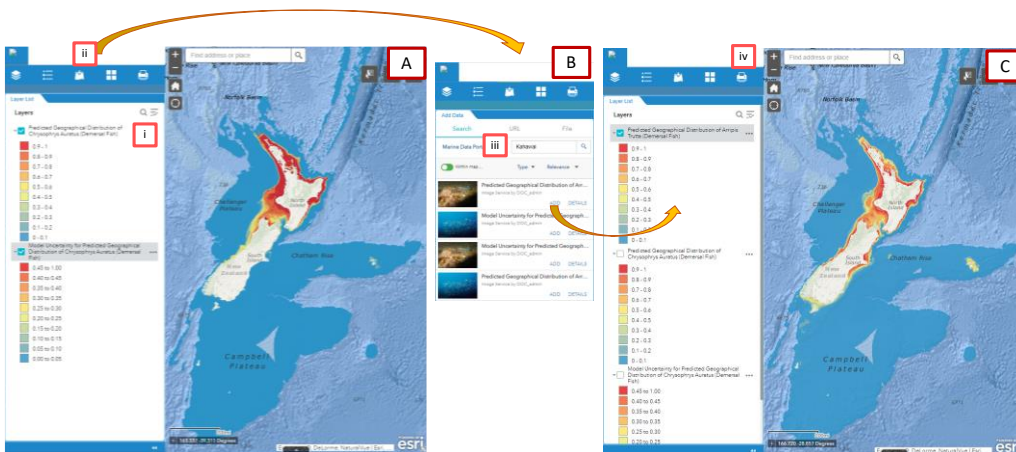
### 3.3 Mapping spatial distributions

Taxa distributions (mean HSI ranging from 0 - 1) and associated uncertainty estimates (measured as the standard deviation of  
the mean HSI) can be mapped within the atlas (Figure S2, A) by clicking the “View on Map” clickable link (Figure S1, ix).  
As with desktop GIS, layers can be displayed or removed by clicking the “tick” box next to the layer names (Figure S2, i), and  
250 the mapped layers can be explored by zooming in or out and focussing on areas of interest by clicking and dragging the mapped  
layer. These features are particularly useful to explore maps for macroalgae and reef fish as these are predicted at a high  
resolution (250 m grid) and exclusively in areas of shallow rocky reefs which have small spatial extents at the scale of the  
Aotearoa New Zealand EEZ/TS.

Multiple taxa distributions and associated uncertainty layers can be added to the map by clicking the “Add Data” tab (Figure  
S2, ii) and using the search function (Figure S2, iii). For example, following a search for “Kahawai” (*Arripis trutta*) (Figure  
S2, B), the layer can be displayed in addition to previously loaded layers (Figure S2, C). The order of the layers and the  
transparency of the layers can be changed to explore overlaps or differences between layers (as per commonly used desktop-  
based GIS software).

### 3.4 Generating maps

260 Maps of predicted distributions and associated uncertainty can easily be generated and exported at various spatial scales using  
the “Print” tab (Figure S2, iv). Options allow customisation of titles, map extent, scale bar units and print quality (e.g., see  
**Error! Reference source not found.** for example output). For more complex display options and further analysis, it is  
recommended that the layers of interest are downloaded and imported to desktop GIS (as described in the metadata section).



265 **Figure S2.** Illustration of mapping features. A) Online GIS for Australasian snapper (*Chrysophrys auratus*); B) Illustration of adding data  
using the search term “Kahawai” (*Arripis trutta*); and C) Online GIS for Australasian snapper and Kahawai.

### References

- Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., and Anderson, R. P.: spThin: an R package for spatial  
thinning of species occurrence records for use in ecological niche models, *Ecography*, 38, 541-545, 2015.
- 270 Allouche, O., Tsoar, A., and Kadmon, R.: Assessing the accuracy of species distribution models: prevalence, kappa and the  
true skill statistic (TSS), *Journal of applied ecology*, 43, 1223-1232, 2006.



- Anderson, A., Stephenson, F., and Behrens, E.: Updated habitat suitability modelling for protected corals in New Zealand waters, NIWA report prepared for Department of Conservation (DOC), NIWA CLIENT REPORT No: 2020174WN 2020.
- 275 Anderson, O. F., Guinotte, J. M., Rowden, A. A., Tracey, D. M., Mackay, K. A., and Clark, M. R.: Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand, *Deep Sea Research Part I: Oceanographic Research Papers*, 115, 265-292, 2016.
- Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W.: Selecting pseudo-absences for species distribution models: how, where and how many?, *Methods in ecology and evolution*, 3, 327-338, 2012.
- 280 Behrenfeld, M. J. and Falkowski, P. G.: Photosynthetic rates derived from satellite-based chlorophyll concentration, *Limnology and oceanography*, 42, 1-20, 1997.
- Bowden, D. A., Anderson, O. F., Rowden, A. A., Stephenson, F., and Clark, M. R.: Assessing Habitat Suitability Models for the Deep Sea: Is Our Ability to Predict the Distributions of Seafloor Fauna Improving?, *Frontiers in Marine Science*, 8, 10.3389/fmars.2021.632389, 2021.
- Breiman, L.: Random forests, *Machine learning*, 45, 5-32, 2001.
- 285 Cael, B., Bisson, K., and Follett, C. L.: Can rates of ocean primary production and biological carbon export be related through their probability distributions?, *Global biogeochemical cycles*, 32, 954-970, 2018.
- CSIRO Atlas of Regional Seas: [www.cmar.csiro.au/cars](http://www.cmar.csiro.au/cars), last access: 14/12/2017.
- Compton, T. J., Bowden, D. A., Roland Pitcher, C., Hewitt, J. E., and Ellis, N.: Biophysical patterns in benthic assemblage composition across contrasting continental margins off New Zealand, *Journal of Biogeography*, 40, 75-89, 2013.
- 290 Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., and Leitão, P. J.: Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, *Ecography*, 36, 27-46, 2013.
- Elith, J., Kearney, M., and Phillips, S.: The art of modelling range-shifting species, *Methods in Ecology and Evolution*, 1, 330-342, 2010.
- 295 Elith, J., Leathwick, J. R., and Hastie, T.: A working guide to boosted regression trees, *Journal of Animal Ecology*, 77, 802-813, doi:10.1111/j.1365-2656.2008.01390.x, 2008.
- Ellis, N., Smith, S. J., and Pitcher, C. R.: Gradient forests: calculating importance gradients on physical predictors, *Ecology*, 93, 156-168, 10.1890/11-0252.1, 2012.
- Folk, R. L., Andrews, P. B., and Lewis, D. W.: Detrital sedimentary rock classification and nomenclature for use in New Zealand, *New Zealand Journal of Geology and Geophysics*, 13, 937-968, 10.1080/00288306.1970.10418211, 1970.
- 300 Frouin, R., McPherson, J., Ueyoshi, K., and Franz, B. A.: A time series of photosynthetically available radiation at the ocean surface from SeaWiFS and MODIS data, *Remote Sensing of the Marine Environment II*, 852519,
- Georgian, S. E., Anderson, O. F., and Rowden, A. A.: 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, 305 <https://doi.org/10.1016/j.fishres.2018.11.020>, 2019.
- Glad, A., Monnet, J.-M., Pagel, J., and Reineking, B.: Importance and effectiveness of correction methods for spatial sampling bias in species with sex-specific habitat preference, *Ecology and evolution*, 9, 13188-13201, 10.1002/ece3.5765, 2019.
- Hijmans, R. J., Phillips, S., Leathwick, J., and Elith, J.: dismo: Species Distribution Modeling R package version 1.1-4. <https://CRAN.R-project.org/package=dismo>, 2017.
- 310 Jouffray, J.-B., Wedding, L. M., Norström, A. V., Donovan, M. K., Williams, G. J., Crowder, L. B., Erickson, A. L., Friedlander, A. M., Graham, N. A., and Gove, J. M.: Parsing human and biophysical drivers of coral reef regimes, *Proceedings of the Royal Society B*, 286, 20182544, 2019.
- Leathwick, J., Elith, J., Francis, M., Hastie, T., and Taylor, P.: 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, 2006.
- 315 Liaw, A. and Wiener, M.: Classification and regression by randomForest, *R news*, 2, 18-22, 2002.
- Lundquist, C., Stephenson, F., McCartain, L., Watson, S., Brough, T., Nelson, W., Neill, K., Anderson, T., Anderson, O., Bulmer, R., Gee, E., Pinkerton, M., Rowden, A., and Thompson, D.: Evaluating Key Ecological Areas datasets for the New Zealand Marine Environment, Project #DOC19206, Hamilton, New Zealand. 120 p., 2020.
- 320 Lutz, M., Dunbar, R., and Caldeira, K.: Regional variability in the vertical flux of particulate organic carbon in the ocean interior, *Global biogeochemical cycles*, 16, 11-11-11-18, 2002.
- Mitchell, J. S., Mackay, K. A., Neil, H. L., Mackay, E. J., Pallentin, A., and Notman, P.: Undersea New Zealand, 1:5,000,000, NIWA Chart, Miscellaneous Series No. 92, 2012.
- Müller, D., Leitão, P. J., and Sikor, T.: Comparing the determinants of cropland abandonment in Albania and Romania using boosted regression trees, *Agricultural Systems*, 117, 66-77, 2013.
- 325 New Zealand fish and squid distributions from research bottom trawls. Southwestern Pacific OBIS, 486781 records: <http://nzobisipt.niwa.co.nz/resource.do?r=obisprovider>, last access: 01/09/2019.
- Catch data from New Zealand research trawls. Southwestern Pacific OBIS, 15157 records: <http://nzobisipt.niwa.co.nz/resource.do?r=trawl>, last access: 01/09/2019.
- 330 Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., and Ferrier, S.: Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data, *Ecological Applications*, 19, 181-197, 2009.
- Pinkerton, M.: Ocean colour satellite observations of phytoplankton in the New Zealand EEZ, 1997–2016, NIWA, Prepared for the Ministry for the Environment. Wellington, 2016.

- 335 Pinkerton, M., Gall, M., Wood, S., and Zeldis, J.: Measuring the effects of bivalve mariculture on water quality in northern New Zealand using 15 years of MODIS-Aqua satellite observations, *Aquaculture Environment Interactions*, 10, 529-545, 2018.
- Pinkerton, M., Rickard, G., Nodder, S., and Macdonald, H.: Climate change impacts and implications: net primary production, particulate flux and impacts on fish species in the New Zealand region, NIWA report 62 pp, 2016.
- 340 Pinkerton, M. H., Décima, M., Kitchener, J. A., Takahashi, K. T., Robinson, K. V., Stewart, R., and Hosie, G. W.: Zooplankton in the Southern Ocean from the continuous plankton recorder: Distributions and long-term change, *Deep Sea Research Part I: Oceanographic Research Papers*, 103303, <https://doi.org/10.1016/j.dsr.2020.103303>, 2020.
- R Core Team: R: A Language and Environment for Statistical Computing [code], 2020.
- Ridgeway, G.: *Generalized Boosted Models: A guide to the gbm package*, 2007.
- 345 Ridgway, K., Dunn, J., and Wilkin, J.: Ocean interpolation by four-dimensional weighted least squares—Application to the waters around Australasia, *Journal of atmospheric and oceanic technology*, 19, 1357-1375, 2002.
- Rowden, A. A., Anderson, O. F., Georgian, S. E., Bowden, D. A., Clark, M. R., Pallentin, A., and Miller, A.: 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, 2017.
- 350 Smith, A. N., Duffy, C., Anthony, J., and Leathwick, J. R.: Predicting the distribution and relative abundance of fishes on shallow subtidal reefs around New Zealand, Department of Conservation, Wellington, New Zealand 2013.
- Stephenson, F., Rowden, A. A., Anderson, O. F., Pitcher, C. R., Pinkerton, M. H., Petersen, G., and Bowden, D. A.: Presence-only habitat suitability models for vulnerable marine ecosystem indicator taxa in the South Pacific have reached their predictive limit, *ICES Journal of Marine Science*, 10.1093/icesjms/fsab162, 2021.
- 355 Stephenson, F., Leathwick, J. R., Geange, S. W., Bulmer, R. H., Hewitt, J. E., Anderson, O. F., Rowden, A. A., and Lundquist, C. J.: Using Gradient Forests to summarize patterns in species turnover across large spatial scales and inform conservation planning, *Diversity and Distributions*, 24, 1641 - 1656, doi:10.1111/ddi.12787, 2018a.
- Stephenson, F., Goetz, K., Sharp, B. R., Mouton, T. L., Beets, F. L., Roberts, J., MacDiarmid, A. B., Constantine, R., and Lundquist, C. J.: Modelling the spatial distribution of cetaceans in New Zealand waters, *Diversity and Distributions*, 26, 495-516, 10.1111/ddi.13035, 2020.
- 360 Stephenson, F., Rowden, A., Anderson, T., Hewitt, J., Costello, M., Pinkerton, M., Morrison, M., Clark, M., Wadhwa, S., Mouton, T., and Lundquist, C.: Mapping Key Ecological Areas in the New Zealand Marine Environment: Data collation, NIWA Client Report 2018332HN, prepared for the Department of Conservation, 2018b.
- Stephenson, F., Rowden, A. A., Brough, T., Petersen, G., Bulmer, R. H., Leathwick, J. R., Lohrer, A. M., Ellis, J. I., Bowden, D. A., Geange, S. W., Funnell, G. A., Freeman, D. J., Tunley, K., Tellier, P., Clark, D. E., Lundquist, C. J., Greenfield, B. L., Tuck, I. D., Mouton, T. L., Neill, K. F., Mackay, K. A., Pinkerton, M. H., Anderson, O. F., Gorman, R. M., Mills, S., Watson, S., Nelson, W. A., and Hewitt, J. E.: Development of a Seafloor Community Classification for the New Zealand Region Using a Gradient Forest Approach, *Frontiers in Marine Science*, 8, 10.3389/fmars.2021.792712, 2022a.
- 370 Stephenson, F., Brough, T., Lohrer, D., Leduc, D., Geange, S., Anderson, O., Bowden, D., Clark, M. R., Davey, N., Pardo, E., Gordon, D. P., Finucci, B., Kelly, M., Macpherson, D., McCartain, L., Mills, S., Neill, K., Nelson, W., Peart, R., Pinkerton, M. W., Read, G., Robertson, J., Rowden, A., Schnabel, K., Stewart, A., Struthers, C., Tait, L., Tracey, D., Weston, S., and Lundquist, C.: An atlas of seabed biodiversity for Aotearoa New Zealand [dataset], <https://doi.org/10.5281/zenodo.7083642>, 2022b.
- Swart, D. H.: Offshore sediment transport and equilibrium beach profiles, Delft Hydraul. Lab. Publ. 131, Delft, 1974.
- 375 Walters, R. A., Goring, D. G., and Bell, R. G.: Ocean tides around New Zealand, *New Zealand Journal of Marine and Freshwater Research*, 35, 567-579, 2001.
- Inherent Optical Properties (IOPs) Algorithm Theoretical Basis Document: <https://oceancolor.gsfc.nasa.gov/atbd/giop/>, last access: 27/07/2020.
- 380 Wright, D., Pendleton, M., Boulware, J., Walbridge, S., Gerlt, B., Eslinger, D., Sampson, D., and Huntley, E.: ArcGIS Benthic Terrain Modeler (BTM), v. 3.0, Environmental Systems Research Institute, NOAA Coastal Services Center, Massachusetts Office of Coastal Zone Management, Available Online at [Http://Esriurl](http://Esriurl). Com, 5754, 2012.