Stephenson & Brough et al., 2023. Supplementary materials 1 – Model fitting and metadata

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), 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 ≥ 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 ≥ 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 ≥ 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 ≥ 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. (2022) (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) collated for species distribution models (table modified from Stephenson et al. (2022).

Abbreviation	Full name	Description	Source
Bathy	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).	Mitchell et al. (2012)
Beddist	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 ≥ 200 m.	Swart (1974); updated in 2019
BotNi	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).	NIWA, unpublished
BotOxy	Dissolved oxygen at depth	Annual average water oxygen concentration at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from Dunn et al. (2002). Oceanographic data from CARS2009 (2009).	NIWA, unpublished
BotPhos	Bottom phosphate	Annual average water phosphate concentration at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from Dunn et al. (2002). Oceanographic data from CARS2009 (2009).	NIWA, unpublished
BotSal	Salinity at depth	Annual average water salinity concentration at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from Dunn et al. (2002). Oceanographic data from CARS2009 (2009).	NIWA, unpublished
BotSil	Bottom silicate	Annual average water silicate concentration at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from Dunn et al. (2002). Oceanographic data from CARS2009 (2009).	NIWA, unpublished
BotTemp	Temperature at depth	Annual average water temperature at the seafloor (using Aotearoa New Zealand bathymetry layer) based on methods from (Ridgway et al., 2002). Oceanographic data from (Cars2009, 2009).	NIWA, unpublished

BPI_broad	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.	NIWA, unpublished
BPI_fine	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.	NIWA, unpublished
Chl a	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).	NIWA unpublished; Based on processing described in Pinkerton et al. (2018) and updated in Pinkerton et al. (2020).
Chl-a.Grad	Chlorophyll-a concentration spatial gradient	Smoothed magnitude of the spatial gradient of annual mean chlorophyl-*a* concentration. Derived from Chl-a described above.	NIWA unpublished, updated in 2020; Based on processing described in (Pinkerton et al., 2018)
DET	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 *adg_443_giop* ocean product (Werdell, 2019).	NIWA unpublished, updated in 2020; Based on processing described in (Pinkerton et al., 2018). Processing for *adg_443_giop* ocean product described in (Werdell, 2019).
Ebed	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.	NIWA unpublished, updated in 2020, based on processing described in Pinkerton et al. (2018)
POCFlux	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. (2017).	NIWA unpublished, updated in 2020. Based on processing described in Pinkerton et al. (2016) with new data from Cael et al. (2018).

PB555nm	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 bbp555 from MODIS-Aqua data. Ocean: *bbp_555_giop* ocean product (Werdell, 2019). Result calculated as long-term (2002–2017) average.	NIWA unpublished, updated in 2020; Based on processing described in Pinkerton et al. (2018). Processing for *bbp_555_giop* ocean product described in Werdell (2019).
SeasTDiff	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.	NIWA, unpublished data, updated in 2018
Sed.class	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.	NIWA unpublished, updated in 2020
Slope	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
SSTGrad	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.	NIWA unpublished, updated in 2020
TC	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*exp(i*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.	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., 2021a).

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., 2021a) 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., 2021a).

Species Distribution Model tuning

To estimate taxonomic distributions, BRT and RF models require locations of both presences (occurrence records) and absences. Here, 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). Target-group background data were generated for each taxon from occurrences within taxonomic groups (i.e., demersal fish relative absences were generated from demersal fish occurrence records). The location of relative absences was required to be at least 1 km from presence data and the number of target-group background 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^2 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 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., 2021a). 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 depths of 2000 m for demersal fish and subtidal invertebrates, 40 m for macroalgae, and were predicted to the full depth range 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. (2021a) and Anderson et al. (2020). This adapted procedure derives a two-part weighting 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:

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

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

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

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

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

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.

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 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 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. 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).

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).

Table S4. Description and evaluation score of expert knowledge of the taxon's distribution

Evaluation score	Description
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.

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

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\Chrysophrys.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 Chrysophrys auratus (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 Chrysophrys auratus (demersal fish) - measured as habitat suitability index (HSI: 0-1).
Description	Detailed description of the layer including important metrics and scores for model evaluation and expert evaluations.	Predicted geographic distribution of *Chrysophrys auratus* (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 Chrysophrys auratus (demersal fish)’).

		Number of taxa records: 2932 Statistical model performance: Good (TSS = 0.88) Expert evaluation of predicted geographical distribution: 1, very accurate Spatial resolution: 1 km
Credits	Credits and reference	This database was prepared to support the Key Ecological Areas (KEA) geospatial database by Fabrice Stephenson and Tom Brough see Stephenson & Brough et al. 2022. 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.
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 *Chrysophrys auratus* (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

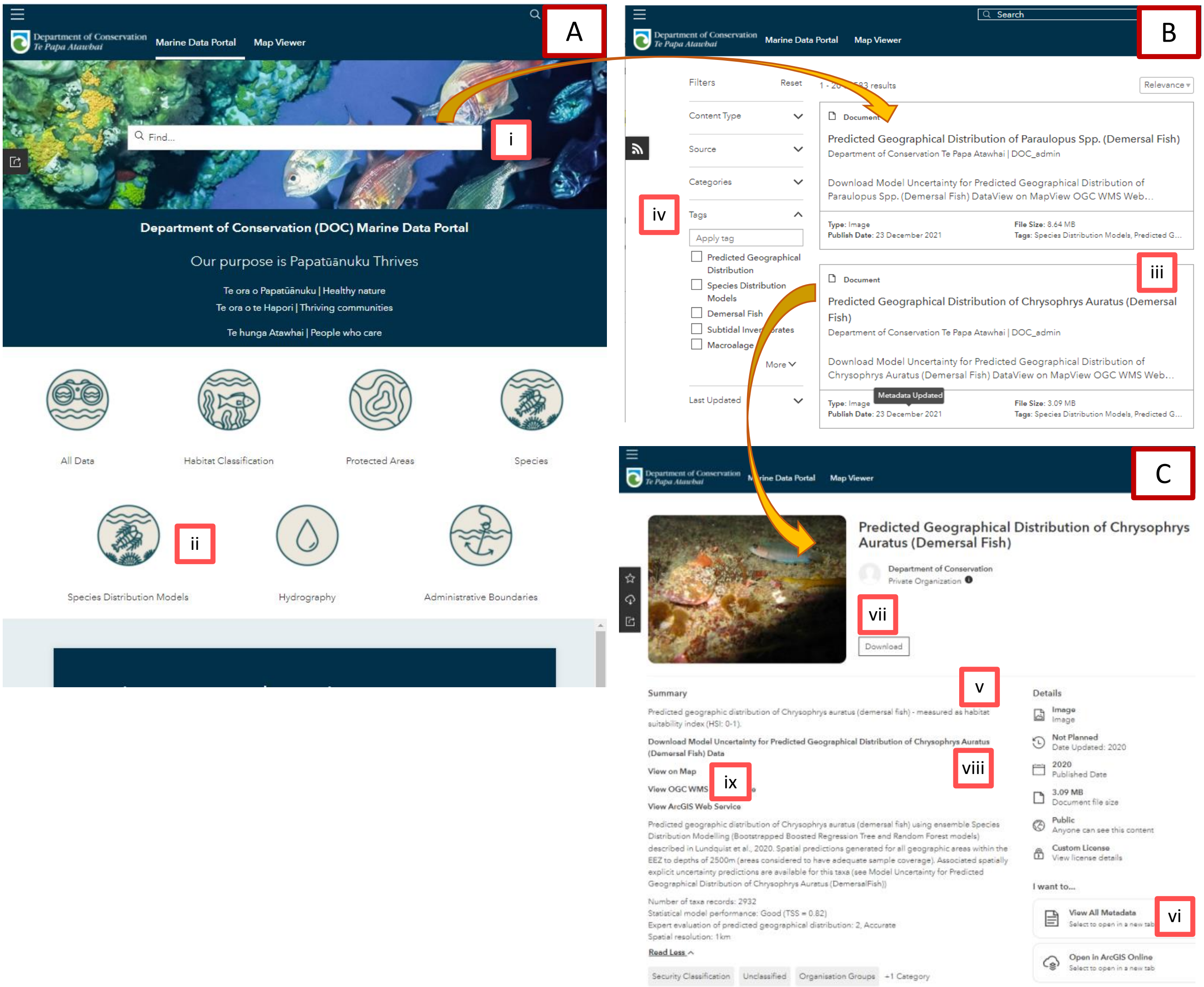
Lineage	Original provenance of the layer	Species occurrence records and environmental co-variates used in the species distribution models were compiled from various open sources and are described in Stephenson et al. (2022). Further information on the expert evaluation process and results is available 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. Marine Ecology Progress Series 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

Stephenson & Brough et al., 2023. Supplementary material 2 – Instructions for the use of the Department of Conservation interactive geoportal

The NZ Department of Conservation online geoportal

This supplement details access, visualisation and download options for the NZ atlas of seafloor biodiversity via an alternative online geoportal housed by the NZ Department of Conservation (https://doc-marine-data-deptconservation.hub.arcgis.com).

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 the available taxa layers (i.e., taxonomic and vernacular taxa names) and statistical and expert evaluation scores see the main access point for the Atlas via the Zonodo database (Stephenson et al., 2022; https://doi.org/10.5281/zenodo.7083642).



Predicted Geographical Distribution of Paraulopus Spp. (Demersal Fish)
Department of Conservation Te Papa Atawhai | DOC_admin
Download Model Uncertainty for Predicted Geographical Distribution of Paraulopus Spp. (Demersal Fish) DataView on MapView OGC WMS Web...
Type: Image
Publish Date: 23 December 2021
File Size: 8.64 MB
Tags: Species Distribution Models, Predicted G...

Predicted Geographical Distribution of Chrysophrys Auratus (Demersal Fish)
Department of Conservation Te Papa Atawhai | DOC_admin
Download Model Uncertainty for Predicted Geographical Distribution of Chrysophrys Auratus (Demersal Fish) DataView on MapView OGC WMS Web...
Type: Image
Publish Date: 23 December 2021
File Size: 3.09 MB
Tags: Species Distribution Models, Predicted G...

Predicted Geographical Distribution of Chrysophrys Auratus (Demersal Fish)
Department of Conservation
Private Organization

Summary
Predicted geographic distribution of Chrysophrys auratus (demersal fish) - measured as habitat suitability index (HSI: 0-1).
Download Model Uncertainty for Predicted Geographical Distribution of Chrysophrys Auratus (Demersal Fish) Data
View on Map
View OGC WMS Service
View ArcGIS Web Service
Predicted geographic distribution of Chrysophrys auratus (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 2500m (areas considered to have adequate sample coverage). Associated spatially explicit uncertainty predictions are available for this taxa (see Model Uncertainty for Predicted Geographical Distribution of Chrysophrys Auratus (DemersalFish))
Number of taxa records: 2932
Statistical model performance: Good (TSS = 0.82)
Expert evaluation of predicted geographical distribution: 2, Accurate
Spatial resolution: 1km
Read Less
Security Classification
Unclassified
Organisation Groups
+1 Category

Details
Image
Image
Not Planned
Date Updated: 2020
2020
Published Date
3.09 MB
Document file size
Public
Anyone can see this content
Custom License
View license details
I want to...
View All Metadata
Select to open in a new tab
Open in ArcGIS Online
Select to open in a new tab

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).

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 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" link in the summary (Figure S1, viii).

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 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).

Generating maps

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 Figure S2 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).

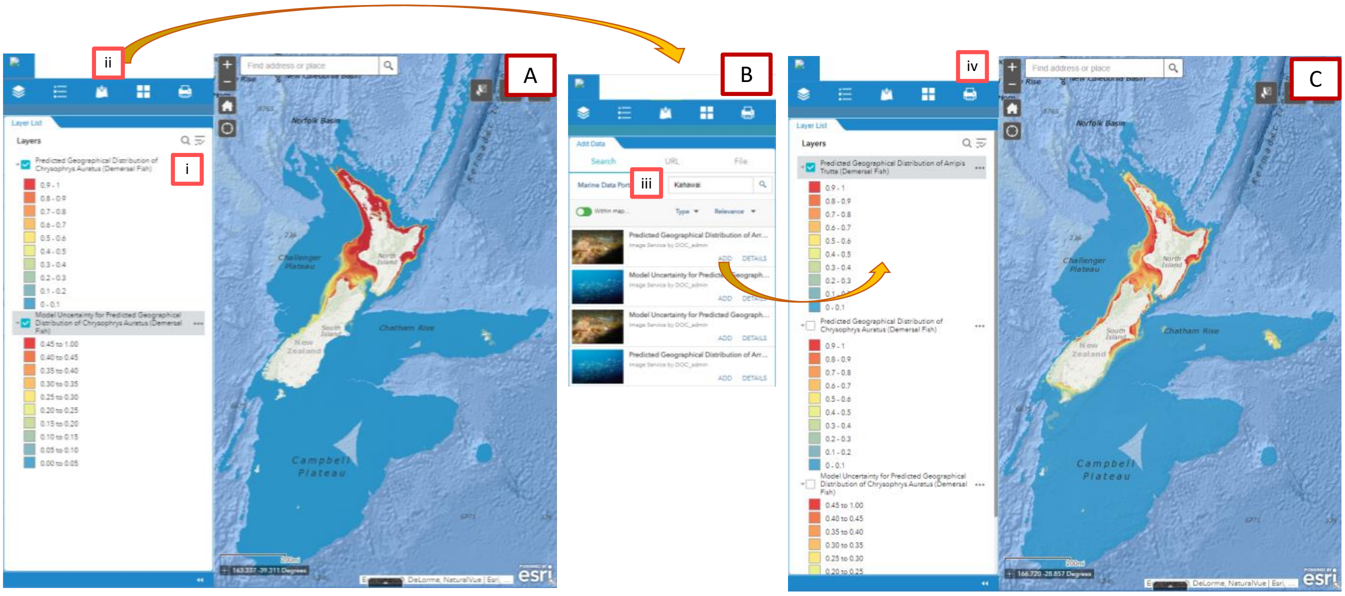


Figure S2. Illustration of mapping features. A) Online GIS for Australasian snapper (*Chrysophrys auratus*); B) Illustration of adding data using the search term “Kahawai” (*Arripis trutta*); and C) Online GIS for Australasian snapper and Kahawai.