BIODIVERSITY RESEARCH

WILEY Diversity and Distributions

Using Gradient Forests to summarize patterns in species turnover across large spatial scales and inform conservation planning

Fabrice Stephenson[1] 🆔 | John R. Leathwick[2] | Shane W. Geange[3] | Richard H. Bulmer[1] | Judi E. Hewitt[1,4] | Owen F. Anderson[5] | Ashley A. Rowden[5] | Carolyn J. Lundquist[1,6]

[1]National Institute of Water and Atmosphere (NIWA), Hamilton, New Zealand

[2]Conservation Science Consultant, Hamilton, New Zealand

[3]Department of Conservation, Wellington, New Zealand

[4]Department of Statistics, University of Auckland, Auckland, New Zealand

[5]National Institute of Water and Atmospheric Research, Wellington, New Zealand

[6]Institute of Marine Science, University of Auckland, Auckland, New Zealand

Correspondence
Fabrice Stephenson, National Institute of Water and Atmosphere (NIWA), Hamilton, New Zealand.
Email: fabrice.stephenson@niwa.co.nz

Abstract

Aim: Producing quantitative descriptions of large-scale biodiversity patterns is challenging, particularly where biological sampling is sparse or inadequate. This issue is particularly problematic in marine environments, where sampling is both difficult and expensive, often resulting in patchy and/or uneven coverage. Here, we evaluate the ability of Gradient Forest (GF) modelling to describe broad-scale marine biodiversity patterns, using a large dataset that also provided opportunity to investigate the effects of sample size on model stability.

Location: New Zealand's Extended Continental Shelf to depths of 2,000 m.

Methods: GF models were used to analyse and predict spatial patterns of demersal fish species turnover (beta diversity) using an extensive demersal fish dataset (>27,000 research trawls) and high-resolution environmental data layers (1 km^2 grid resolution). GF models were fitted using various sized, mutually exclusive subsets of the demersal fish data to explore the effect of variation in numbers of training observations on model performance and stability. A final GF model using 13,917 samples was used to transform the environmental layers, which were then classified to produce 30 spatial groups; the ability of these groups to identify fish samples with similar composition was evaluated using independent sample data.

Results: Model fitting using varying sized subsets of the data indicated only minimal changes in model outcomes when using >7,000 observations. A multiscale spatial classification of marine environments created using results from a final GF model fitted using ~14,000 samples was highly effective at summarizing spatial variation in both fish assemblage composition and species turnover.

Main conclusions: The hierarchical nature of the classification supports its use at varying levels of classification detail, which is advantageous for conservation planning at differing spatial scales. This approach also facilitates the incorporation of information on intergroup similarities into conservation planning, allowing greater protection of distinctive groups likely to support unusual assemblages of species.

KEYWORDS
beta diversity, biodiversity, Gradient Forest models, marine, predictive models, sample size, spatial patterns, species turnover

1 | INTRODUCTION

Sound conservation planning and management at the ecosystem level depend heavily on the reliable characterization of biodiversity patterns, often across large regions (Ferrier, Manion, Elith, & Richardson, 2007; Margules & Pressey, 2000). Ideally, this characterization is based on quantitative information describing different components of diversity, including not only the distributions of individual species and/or communities, but also of emergent properties such as alpha diversity (local richness) and beta diversity (species turnover along spatial or environmental gradients; Harrison, Ross, & Lawton, 1992; Legendre, Borcard, & Peres-Neto, 2005; Nekola & White, 2002; Shmida & Wilson, 1985). These latter two components are particularly valuable for spatial planning; while the total number of species (alpha diversity) contributes to the relative importance of an area for conservation, it is the rate of species turnover between sites that largely determines the optimal spatial arrangement of conservation areas (Arponen, Moilanen, & Ferrier, 2008; Bush, Harwood, Hoskins, Mokany, & Ferrier, 2016; McKnight et al., 2007; Nekola & White, 1999, 2002; Socolar, Gilroy, Kunin, & Edwards, 2016).

Unfortunately, producing quantitative descriptions of large-scale biodiversity patterns is challenging, particularly where biological sampling is sparse or inadequate (Elith et al., 2006; Leaper et al., 2011; Morán-Ordóñez, Lahoz-Monfort, Elith, & Wintle, 2017). It is often particularly problematic in marine environments, where sampling is both difficult and expensive, resulting in patchy and/or irregular coverage (Leathwick, Elith, Francis, Hastie, & Taylor, 2006; Thomson et al., 2014). Spatial classifications of environments are one tool commonly used to overcome these difficulties, particularly at larger spatial scales (Snelder et al. 2007). Such classifications typically use more readily available environmental data to group together sites likely to have similar biological character. These classifications can be used to identify areas that are likely to respond in similar ways to human activity or management actions (Bailey, 1983), to define standards for the management of human impacts (Omernik & Bailey, 1997), to stratify sites for surveying and monitoring programs (Hawkins et al., 2000) or to identify priority areas for conservation management (Leathwick et al., 2012). Earlier classification methods relied largely on subjective, expert decision-making (e.g., thematic or hierarchical classifications, Roff, Taylor, and Laughren (2003)), while a number of recent classifications use quantitative approaches, sometimes with manually imposed rules (Snelder and Biggs, 2002) or, alternatively, use numerical classification procedures to group sites (Belbin, 1993). In addition, qualitative classifications often lack information on within-class variation in the community composition (Leaper et al., 2011), instead relying on subjective alignment with coarse categorical habitat features, that is, seagrass beds, rocky reef, sand and mud (Mumby et al., 2008). Regardless, subjective decisions are required in the choice of environmental factors, how they should be weighted and whether transformations should be applied to increase their ability to distinguish biodiversity patterns (Leathwick et al., 2011).

One approach to overcoming this difficulty is to combine continuous environmental data with biological samples (Anderson et al., 2016; Dunstan, Althaus, Williams, & Bax, 2012), using analytical tools that quantitatively assess the role that different environmental factors play in influencing biodiversity patterns (Ferrier et al., 2004; Pitcher et al., 2012). Results are then used to control the selection, weighting and transformation of environmental predictors to be classified, increasing the ability of the classification groups to represent spatial variation in biodiversity character. Generalized dissimilarity modelling (GDM—Ferrier et al., 2007; Ferrier, 2002) was one of the first techniques to explicitly model relationships between environment and species turnover. It uses a generalized linear modelling framework to identify transformations of the environmental predictors that maximize their ability to predict biological distances (Bray–Curtis similarity) between sample sites. The fitted transformations can then be used to predict community turnover across the entire study region, including within areas lacking biological samples (Ferrier et al., 2007).

A second technique, Gradient Forest (GF; Pitcher, Ellis, and Smith, 2011) is a more recent development that produces similar outputs to GDM, but using a fundamentally different approach (Leaper et al., 2011). While GDM fits a single regression model that describes the relationship between environment and species turnover, a GF model consists of an aggregation of Random Forest (RF; Breiman, 2001) models, each of which describes the environmental relationships of an individual species. Once these models have been fitted, the information that they contain about the relative importance of different predictors, and where changes in the presence (or abundance) of the modelled species occur along each of their ranges, is aggregated and used to build transforms that maximize for each predictor their correspondence with species turnover (Pitcher et al., 2011). These transforms closely resemble those produced by GDM and can be used in similar ways, for example, these values can be calculated across extensive geographic areas, with the transformed predictors then classified to define spatial groups that capture variation in species composition and turnover, making them potentially well suited as input to conservation planning analyses. The more incremental approach to model fitting in GF makes it particularly well suited to the analysis of large datasets, whose size can be limiting in GDM.

Here, we evaluate the ability of GF to describe broad-scale biodiversity patterns for conservation planning. The analysis combines high-resolution environmental data (1 km^2 grid resolution) across New Zealand's Extended Continental Shelf to depths of 2,000 m and distributional data for 253 demersal fish species from c. 27,000 research trawls. We begin by exploring relationships between the size of the training dataset and model stability, using results to identify a conservatively generous training dataset with which to fit a final model. Results from this final model are used to transform the environmental layers to maximize their correspondence with species turnover, with numerical classification of these transformed layers then used to define spatial groups having similar species composition. Finally, we assess the ability of this classification to represent variation in both species composition and turnover, crucial "success"
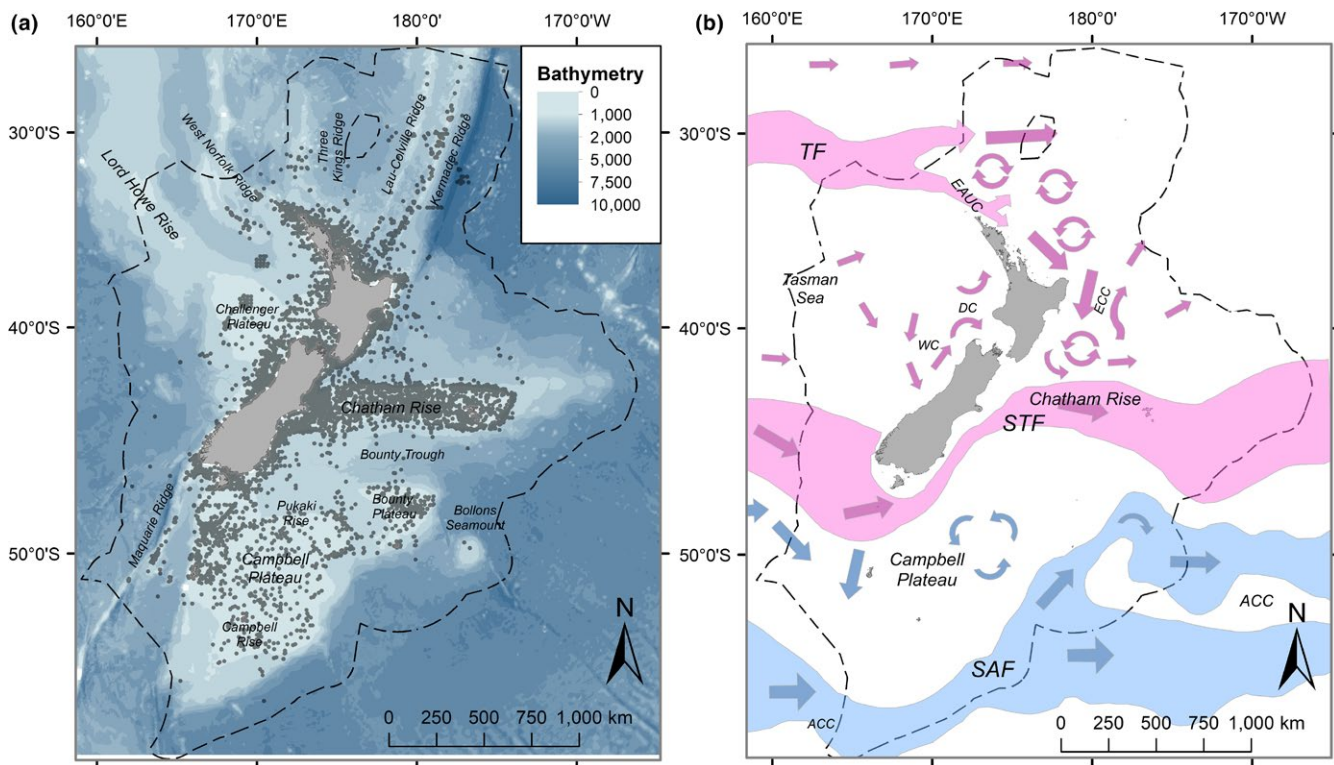
FIGURE 1 Maps of the study region (New Zealand Extended Continental Shelf, black dashed line) showing: (a) bathymetry, sample locations (grey dots) and feature names used in the text; (b) approximate positions and direction of travel of the Tasman Front (TF and its associated currents: the east Auckland Current (EAUC) and East Cape Current (ECC) in the north-east, and the Westland Current (WC) and D'Urville Current (DC) in the West of the study area), Subtropical Front (STF) and Subantarctic Front (SAF and the Antarctic Circumpolar Current (ACC)). Adapted from Carter (2001) [Colour figure can be viewed at wileyonlinelibrary.com]

factors if the classification is to have utility for systematic broad-scale conservation planning.

2 | METHODS

2.1 | Study area

The study area consisted of those parts of the New Zealand Extended Continental Shelf (Figure 1) with depths in the range 0–2,000 m; including deeper regions would have been desirable, but fish samples from sites deeper than 2,000 m are limited (Figure 1a). New Zealand has a long and narrow land mass extending across a wide latitudinal range (≈35–48°S), resulting in a diverse range of environmental conditions in its surrounding waters (Bradford-Grieve, Lewis, & Stanton, 1991; Leathwick et al., 2006). The dominant oceanographic feature is the Subtropical Front (STF), a highly productive zone of mixing between high salinity, nutrient-poor, warm, northern waters and low salinity, nutrient-rich, cold, southern waters (Figure 1b). Currents flow in an eastward direction along the STF but are deflected southwards around the lower South Island, returning north along the east coast of the South Island to resume their eastwards flow along the Chatham Rise (Figure 1b; Bradford-Grieve et al., 2006). Several gyres occur within this mixing zone, mostly to the east of the North Island

(Figure 1b; Bradford-Grieve et al., 2006). Although the waters surrounding New Zealand do not have large areas of continental shelf, extensive submarine plateaus are located to the east, south and west: the Chatham Rise to the east (Leathwick et al., 2006) (Figure 1a); the Bounty Plateau, Pukaki Rise and Campbell Plateau and Rise to the south and south-east (Figure 1a); and the Lord Howe Rise and Challenger Plateau to the west (Figure 1a). Deeper abyssal waters occur in the south-west of the study area along the Puysegur Trench and in the north-east along the Kermadec Trench (Figure 1a).

2.2 | Biological data

Fish are the most abundant and diverse group of vertebrate animals on earth and play a key role in structuring marine ecosystems; interactions within and between fish species and their physical environment are important in defining community structure, diversity and stability in marine ecosystems (Francis, Hurst, McArdle, Bagley, & Anderson, 2002).

Demersal fish abundance data used for this analysis were collected during research trawl surveys conducted between 1979 and 2005; the majority of these data are available from the NIWA Environmental Information Browser (https://ei.niwa.co.nz/; Figure 1). To minimize the effects of variation in individual species

TABLE 1 Environmental variables used as predictors in Gradient Forest analyses

Abbreviation	Full name	Description	Original resolution	Units	Source
Bathy	Bathymetry	Depth at the seafloor was interpolated from contours generated from various sources, including multibeam and single-beam echo sounders, satellite gravimetric inversion and others (Mitchell et al., 2012)	250 m	m	CANZ (2008)
Beddist	Benthic sediment disturbance	Combination of seabed orbital velocities (estimates the average mixing at the seafloor as a consequence of orbital wave action, calculated from a wave climatology-derived hindcast (1979–1998) of swell wave conditions in the New Zealand (NZ) region; Gorman, Bryan, & Laing, 2003) and friction velocity for seabed types (based on grain size). Benthic sediment disturbance from wave action was assumed to be zero where depth ≥200 m	1 km	Unitless	NIWA, unpublished
BotTemp	Temperature at depth	Annual average water temperature at the seafloor (using NZ bathymetry layer) based on methods from Ridgway, Dunn, and Wilkin (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)	250 m	°C/km	NIWA, unpublished
BotNi	Bottom nitrate	Annual average water nitrate concentration at the seafloor (using NZ bathymetry layer) based on methods from Ridgway et al. (2002). Oceanographic data from CARS2009 (2009)	250 m	µmol/L	NIWA, unpublished
BotOxy	Dissolved oxygen at depth	Annual average water dissolved oxygen concentration at the seafloor (using NZ bathymetry layer) based on methods from Ridgway et al. (2002). Oceanographic data from CARS2009 (2009)	250 m	ml/L	NIWA, unpublished
BotSal	Salinity at depth	Annual average water salinity concentration at the seafloor (using NZ bathymetry layer) based on methods from Ridgway et al. (2002). Oceanographic data from CARS2009 (2009)	250 m	Psu	NIWA, unpublished
BotSil	Bottom silicate	Annual average water silicate concentration at the seafloor (using NZ bathymetry layer) based on methods from Ridgway et al. (2002). Oceanographic data from CARS2009 (2009)	250 m	µmol/L	NIWA, unpublished
Disorgm	Coloured dissolved organic matter (CDOM)	Indicative of coloured dissolved organic matter (CDOM) absorption at 440 nm. Based on SeaWiFS ocean colour remote sensing data; modified Case 2 atmospheric correction; modified Case 2 inherent optical property algorithm (Pinkerton et al., 2005)	4 km	Indicative of CDOM absorption at 440 nm $a_g(440)\ (m^{-1})$	Pinkerton (2016)
Roughness	Roughness	Roughness of the seafloor calculated as the standard deviation of depths in a surrounding 3 × 3 km neighbourhood (Leathwick et al., 2012)	250 m	Unitless	Leathwick et al. (2012)NIWA, unpublished data
SeasTDiff	Annual amplitude of seafloor temperature	Smoothed difference in seafloor temperature between the three warmest and coldest months. Providing a measure of temperature amplitude through the year	250 m	°C/km	NIWA, unpublished data

(Continues)

TABLE 1 (Continued)

Abbreviation	Full name	Description	Original resolution	Units	Source
Sed	Sediment type	Seabed sediment and rock data which were obtained from research surveys around the NZ region (≈6,000 data points) were combined with sediment type data from a global online database (Jenkins, 2010), providing a total of about 30,000 data points for the study area. Because of the uneven distribution of these data, interpolation was required to provide a value for each cell. Interpolation was carried out using a kriging process in ArcGIS 10.3.1 (ESRI, 2015) providing coarse categorical classification of sediment type around NZ	polygon data—compiled from maps at various scales	(1) calc-gravel, (2) calc-mud, (3) calc-sand, (4) Clay, (5) deep ocean clays, (6) gravel, (7) mud, (8) sand, (9) siliceous ooze and (10) volcanic	Anderson et al., 2016; NIWA, unpublished data
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 Sea-Viewing Wide Field-of-view Sensor (SeaWiFS) satellite imagery (Pinkerton et al., 2005)	1 km	°C/km	
SuspPM	Suspended particulate matter	Indicative of total suspended particulate matter concentration. Based on SeaWiFS ocean colour remote sensing data (Pinkerton & Richardson 2005); modified Case 2 atmospheric correction; modified Case 2 inherent optical property algorithm (Pinkerton et al. 2006)	4 km	Indicative of total suspended particulate matter concentration (g/m³)	Pinkerton (2016)
TidalCurr	Tidal current speed	Maximum depth-averaged (NZ bathymetry) flows from tidal currents calculated from a tidal model for New Zealand waters (Walters, Goring, & Bell, 2001)	250 m	m/s	NIWA, unpublished data
VGPM	Productivity Model	Provides estimates of surface water primary productivity based on the vertically generalized productivity model of Behrenfeld and Falkowski (1997). Net primary productivity by phytoplankton (mean daily rate of water column carbon fixation) is estimated as a function of remotely sensed chlorophyll concentration, irradiance and photosynthetic efficiency estimated from remotely sensed Sea-Viewing Wide Field-of-view Sensor (SeaWiFS) satellite imagery (M. Pinkerton, NIWA, pers. Comm.)	9 km	mgC m^{-2} day^{-1}	NIWA, unpublished
*DynOc**	Dynamic oceanography	Mean of the 1993–1999 period sea surface above geoid	250 m	m	NIWA, unpublished
*BotOxySat**	Oxygen saturation at depth	Annual average oxygen saturation at the depths	250 m	µmol/L	NIWA, unpublished
*OxyUt**	Apparent oxygen utilization	The difference between the measured dissolved oxygen concentration and its equilibrium saturation concentration in water with the same physical and chemical properties	250 m	µmol/L	NIWA, unpublished

Note. Variables not used in the final analyses because of their high correlation with other variables are identified with an asterisk.

catchability among surveys, especially given the sometimes incomplete documentation of trawl parameters, species catch information was reduced to a presence/absence level (Hewitt, Wang, Francis, Lundquist, & Duffy, 2015; Leathwick et al., 2006). The final dataset included observations of 253 species at 27,440 sample locations (Figure 1a). Some geographic bias is evident in the distribution of samples, with many trawls made at 43–45°S, reflecting the high priority given to surveying commercially important species, that is, hoki (*Macruronus novaezelandiae*), orange roughy (*Hoplostethus atlanticus*) and oreos (mainly *Allocyttus niger* and *Pseudocyttus maculatus*) (Francis et al., 2002) on the Challenger Plateau and Chatham Rise. In some areas, difficult terrain resulted in a paucity of tows, for example, the narrow and topographically complex continental shelf off Fiordland (on the West coast at 45–46°S) and the Kermadec and Colville Ridges (Figure 1a; Francis et al., 2002). This sampling bias was not specifically addressed here as it was assumed that the substantial number of observations across a broad range of environmental gradients provided adequate coverage for species distribution modelling.

2.3 | Environmental data

To capture variability in the marine environments surrounding New Zealand, eighteen high-resolution gridded environmental predictors, mostly at a native resolution of 250 m, were collated and imported into ArcGIS (version 10.4). These variables were selected based on their known influence on demersal fish settlement, growth, survival and distribution (Beentjes, Bull, Hurst, & Bagley, 2002; Bull, Livingston, Hurst, & Bagley, 2001; Francis et al., 2002; Kendrick & Francis, 2002; Leathwick et al., 2006, 2012) and consequent likely influence on fish species assemblage composition and turnover (Table 1). For example, several physicochemical variables are likely to be critical physical determinants of habitat suitability for most fish species, including annual averages of seafloor temperature (*BotTemp*), salinity (*BotSal*), nitrate (*BotNi*), oxygen (*BotOxy*) and silicate (*BotSil*); benthic sediment disturbance (*Beddist*), which is an important feature in shallower depths (<200 m); ocean productivity, as described by vertically generalized production model (*VGPM*), with complementary information on surface water productivity provided by a spatial summary of gradients in sea surface temperature (*SstGrad*); other variables that act as surrogates for a range of correlated biophysical variables, for example, Depth (*Bathy*), seafloor roughness (*Roughness*) and sediment type (*Sed*). Although most of the chosen ocean climate variables (e.g., seafloor temperature, *BotTemp*) were formulated as mean annual statistics, one variable described the annual range in temperature (e.g., the annual temperature range, *SeasTDiff*), showing greatest variability in inshore waters.

Several predictors showed strong collinearity, for example, between depth and seafloor temperature, salinity and oxygen saturation (Ellis, Smith, & Pitcher, 2012). Although GF modelling is reasonably robust to correlated variables due to the incorporation of a conditional permutation approach in the calculation of predictor importance (for further information see Ellis et al. (2012)), the use of highly correlated variables generally provides only minimal improvement in prediction accuracy and complicates interpretation of model outcomes. Consequently, three of the most highly correlated variables were excluded: dynamic oceanography, oxygen saturation at depth and apparent oxygen utilization (Table 1). The remaining 15 variables were retained for all models, ignoring correlations, as all were thought to likely be of ecophysiological importance (see Supporting information Figure S1 for correlations of the final variables included in the model and Table 1 for description of variables). Prior to model fitting, values for each environmental variable were derived for all trawl locations by overlaying these onto each of the environmental predictor layers using the raster package in R (Hijmans & van Etten, 2012).

2.4 | Model fitting and evaluation

GF modelling has two components: the production of a random forest model (Breiman, 2001) for each of the input species (using an extended modelling procedure in R package "extendedForest"; Liaw & Wiener, 2002) and the collation of all the individual split points from these models to calculate species turnover along each environmental gradient (using the R package "gradientForest"; Ellis et al., 2012). All analyses were undertaken in the statistical computing software R (R Core Team, 2013).

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). The predictive power of the individual Random Forest models is evaluated using a measure of R^2_f for each species f (the proportion of out-of-bag variance explained; Ellis et al., 2012). The importance of each predictor variable (measured as R^2) in the model is assessed by quantifying the degradation in performance when each predictor variable is randomly permuted (Pitcher et al., 2012) using a conditional approach which accounts for collinearity between predictor variables (Ellis et al., 2012). GF aggregates the values of the tree splits from the Random Forest models for all species' models with positive fits ($R^2_f > 0$) to develop empirical distributions that represent species turnover along each environmental gradient (Compton, Bowden, Pitcher, Hewitt, & Ellis, 2013; Ellis et al., 2012; Pitcher et al., 2012). The turnover function is measured in dimensionless R^2 units where species with highly predictive random forest models (high R^2_f values) have greater influence on the turnover functions than those with low predictive power (lower R^2_f). The shapes of these turnover curves describe the rate of compositional change along each environmental predictor; steep parts of the curve indicate fast assemblage turnover, and flatter parts of the curve indicate more homogenous regions (Ellis et al., 2012; Pitcher et al., 2012).

In an initial set of analyses, we fitted GF models using subsamples of the full dataset to investigate the relationship between sample size and model performance and stability as measured by species predictive performance (R^2_f), environmental predictor contributions

(R^2) and predictor responses. Separate GF models were fitted to five mutually exclusive subsets of 1,000, 2,000, 3,000 and 5,000 observations and four mutually exclusive subsets of 6,860 observations, all of which were randomly selected from the full dataset. Although distributional data for inadequately sampled species (e.g., <30 occurrences) are generally considered as unsuitable for fitting individual species distribution models (SDMs; Leathwick et al. 2006, Hewitt et al. 2015), here all species with more than 10 observations were included in the analyses. This is because the GF model consists of a set of Random Forest models weighted by their goodness-of-fit; as a consequence, any species model with a positive R^2_f is able to contribute to the combined estimate of species turnover, while those with no predictive power are automatically discarded.

Based on results from these analyses, the full dataset was split into two randomly selected, mutually exclusive subsets, each containing 13,917 observations; a final GF model was fitted to the first of these subsets, while the second subset was used only for model evaluation as described below. The species turnover functions produced by the final GF model were used to create a transformed set of environmental predictor layers (using the predict function; Pitcher et al., 2011), with values in these layers representing species turnover along the range of each environmental predictor.

This set of transformed environmental layers was further analysed in two ways. First, variation within the transformed environmental layers was summarized using principal component analysis (PCA; Pitcher et al., 2011) to provide a generalized description of relationships between species turnover and environment. Second, the transformed environmental dataset was classified in two stages in R (Leathwick et al. 2011) to produce a classification map more suitable for conservation planning purposes. In the first stage, the 2.48 million data points were clustered to form 500 initial groups using nonhierarchical, k-medoids clustering with the Manhattan distance metric as implemented in the function clara in the R package "cluster" (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2017). To allow the classification to be used at varying levels of detail, relationships between these 500 groups were then summarized by agglomerative clustering using flexible UPGMA, the Manhattan metric and a value for beta of −0.1 (Belbin, Faith, & Milligan, 1992), as implemented in the function agnes in the R package "cluster" (Maechler et al., 2017). Results from the classification analysis were mapped geographically to allow inspection of outcomes at varying degrees of classification detail through the range from 10 to 100 groups in steps of ten. We describe results here at a 30-group level of classification to simplify presentation, although a higher level of classification could be more appropriately applied in conservation planning, particularly for inshore waters where greater variation occurs at finer spatial scales (Leathwick et al., 2012; Snelder et al., 2007; see Section 4).

The ability of the transformed environmental layers, classified at a 30-group level, to summarize species composition and turnover as recorded in the biological data was evaluated by spatial overlay of the location of each biological sample in the training dataset onto a digital map of the classification, tagging them with the classification group occurring there. A matrix was then constructed describing the average species composition of observations in each classification group, and this was used to calculate extended biological intergroup distances using the Bray–Curtis distance measure with the functions distance and stepacross implemented in the R package "ecodist" (Goslee & Urban, 2007). Correlations were then calculated between these biological distances and the equivalent intergroup distances (Manhattan metric) from the classification of the transformed environmental layers; because of the noise inherent in the individual species observations, we report these correlations for those groups with at least 100 and 200 biological observations, that is, sufficient to derive a reliable estimate of average species composition. Finally, we repeated this process using the independent evaluation dataset of 13,917 observations to test the utility of the classification to summarize species turnover in a completely independent set of samples.

3 | RESULTS

3.1 | Effects of variation in numbers of training observations on Gradient Forest models

The number of species effectively modelled (i.e., with an $R^2_f > 0$), their predictive performance (R^2_f) and the relative contributions of environmental predictors (R^2) were all sensitive to the number of observations used for model fitting. On average, only 202 species out of 253 were effectively modelled when 1,000 observations were used for model fitting (Figure 2a), but this number increased steadily with increasing dataset size, gradually plateauing off to a mean of 249 species in the four models fitted with 6,860 observations.

The mean predictive performance also increased with increasing sample size but only across those 85 species that were effectively modelled in all five of the initial models fitted to the smallest dataset (1,000 observations) and in all subsequent models (dark grey—Figure 2b). There was a progressive and significant increase in predictive performance for these species up to about 5,000 observations (albeit these mean increases in predictive performance were small, range: 0.51–0.54); only muted further increases occurred beyond this. By contrast, when the mean predictive performance was calculated across all species effectively modelled in each model from the different sized datasets, there was no increase in predictive performance with increasing dataset size (light grey bars—Figure 2b); this largely reflected the increased inclusion of species with few observations as the size of the training dataset was increased, with models for these species generally having lower predictive performance.

Marked differences were also apparent in the relative contributions of the different environmental predictors across the models fitted with datasets of varying size (Figure 3). In GF models with low sample number (1,000 samples), bathymetry (*Bathy*) was the most influential environmental predictor in the analysis, followed closely by dissolved oxygen at depth (*BotOxy*), tidal current speed (*TidalCurr*), temperature at depth (*BotTemp*) and salinity at depth (*BotSal*; Figure 3). While these same environmental predictors continued to play a dominant role as the number of observations
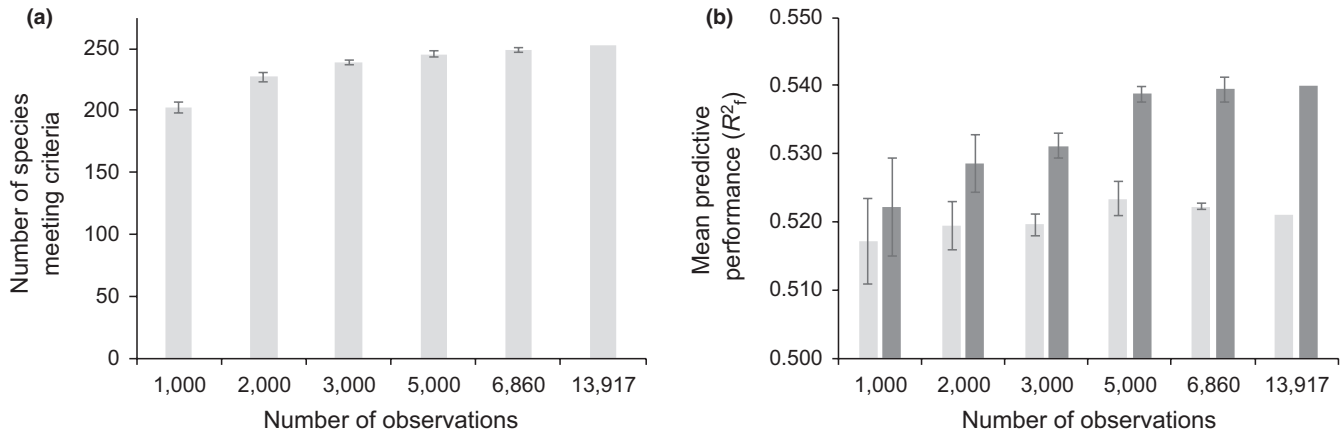
(a)

(b)

FIGURE 2 Numbers of species effectively modelled and their predictive performance as a function of training data sample size—separate Gradient Forest models were fitted using five independent samples of 1,000 through 5,000 observations and four samples of 6,860 independent observations; results from a single final model using 13,917 observations are included for comparison. (a) Mean numbers of fish species effectively modelled versus sample size. (b) Mean predictive performance (R^2_f) for all demersal fish species included in each GF analyses (light grey bars) and mean predictive performance (R^2_f) for fish species successfully fitted in all 24 models (n = 85—dark grey bars) versus sample size. Error bars in all plots show the standard error of the mean

used for model fitting increased, their order of importance differed (Figure 3). Although *Bathy* showed a small increase in its contribution in GF models fitted with larger numbers of samples (6,860 observations), the contributions of *BotOxy*, *TidalCurr*, *BotTemp* and *BotSal*, each became more important than depth with larger sample sizes. Importantly, the stability of these contributions also increased as sample size increased (i.e., standard errors of their means decreased).

3.2 | The final model Gradient Forest model

The final GF model was fitted using all 13,917 observations in our training dataset, assuming that further increasing the number of observations was unlikely to yield any substantial increase in predictive performance; all available species (n = 253) had an R^2_f greater than zero in this model. Species performance in this model (R^2_f) averaged 0.521, with a range from 0.300 (*Eurypharynx pelecanoides*—pelican eel) to 0.913 (*Hoplostethus atlanticus*—orange roughy); there was a weak but positive curvilinear relationship between numbers of positive observations for species and their predictive performance (Figure 4). The five species with the lowest R^2_f values (0.3–0.4) had a mean of 125 positive observations, while the five species with the highest R^2_f values (0.8–0.913) had a mean of 4,558 positive observations (see Supporting information Table S1 for a full list of species, number of positive observations and their associated R^2_f values).

Environmental transformations from the final GF model, constructed using the split information contained in the individual species models, indicate both the overall influence of each environmental predictor and the cumulative changes in species turnover along its range (Figure 5), while the ranges of the fitted functions indicate the relative amounts of species turnover associated with each predictor. Steep parts of the curve indicate fast assemblage turnover, and flatter parts of the curve indicate more homogenous
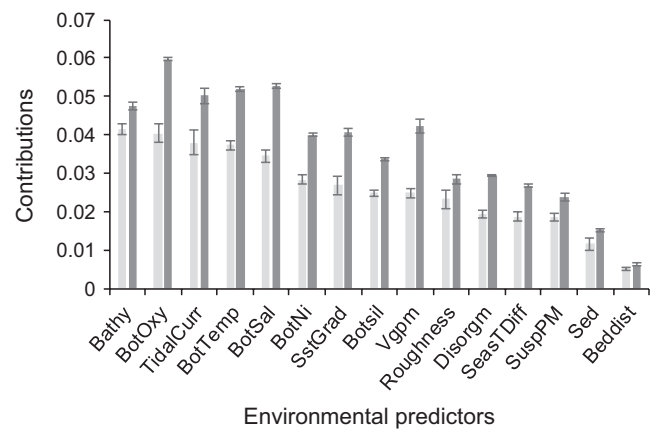

FIGURE 3 Relative importance of environmental variables for predicting demersal fish species turnover as assessed by Gradient Forest models; bars show the mean contribution of each predictor, averaged across five models, and using 1,000 observations (light grey; n = 5) and 6,860 observations (dark grey; n = 4). Error bars show the standard error of the mean

regions (Ellis et al., 2012; Pitcher et al., 2012). Greatest species turnover was associated with the predictor describing dissolved oxygen concentrations at depth (*BotOxy*) (maximum cumulative importance: 0.06), followed closely by salinity at depth (*BotSal*), tidal current speed (*TidalCurr*) and temperature at depth (*BotTemp*), with maximum values for these ranging from 0.04 to 0.05 (Figure 5). Turnover in relation to other environmental predictors was generally lower (ranging from 0.005 to 0.04). Inspection of the fitted functions for these predictors indicate that most had broadly linear relationships with species turnover, although a number showed a levelling off in species turnover at higher values, for example, at depths >1,500 m, tidal current speeds >1.0 m/s and dissolved organic matter >0.6 mgC m^{-2} day^{-1}.

3.3 | Classification of the transformed environmental layers

A generalized model of fish species turnover in New Zealand's Extended Continental Shelf to depths of 2,000 m was produced using GF (Supporting information Figure S2). Although the classification of the transformed environmental layers is capable of being
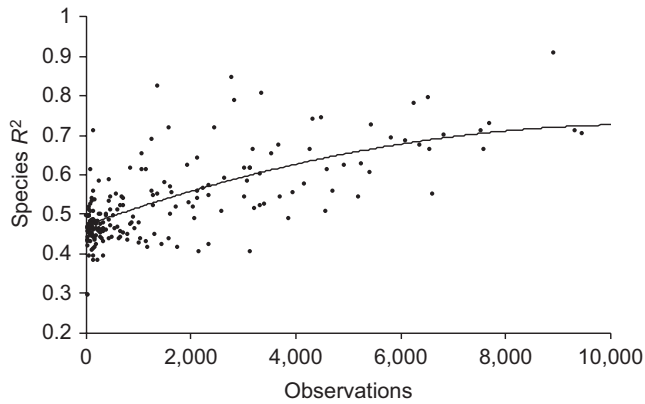

FIGURE 4 Individual species R^2 versus numbers of positive observations for the final Gradient Forest model using 13,917 observations. The black line represents a line of best fit for illustrative purposes

viewed at widely varying levels of detail, for parsimony, we describe thirty groups here (Figure 6). These vary widely not only with respect to the transformed environmental predictors (Figure 7a), but also in their geographic distributions and extents (Figure 6b,c), and their physical and biological characteristics, as described by the raw environmental predictors and species distribution data (Supporting information Table S2).

Broadly, environmental differences between classification groups were relatively muted in the deepest waters (e.g., groups 1, 2 and 13 and groups 3–6, 10 left of Figure 7a, top Figure 6), but differences in temperature, oxygen and salinity become increasingly important in intermediate and shallow depths, for example, groups 7, 9 and 14 occur in more saline, less oxygenated waters (centre of both Figures 6 and 7a) and groups 8, 20, 26 and 28–30 occur in less saline, more oxygenated waters (centre of Figure 6 and lower middle of Figure 7a). Variation in productivity, sea surface temperature gradients, tidal currents, suspended sediments and dissolved organic matter were important differentiating factors in shallow, inshore waters (lower Figure 6 and right of Figure 7a). A more detailed description of the geographic distributions, extents and physical and biological characteristics of the 30-group classification and a larger scale map of the classified environmental layers is contained in the Supporting information Figure S3.
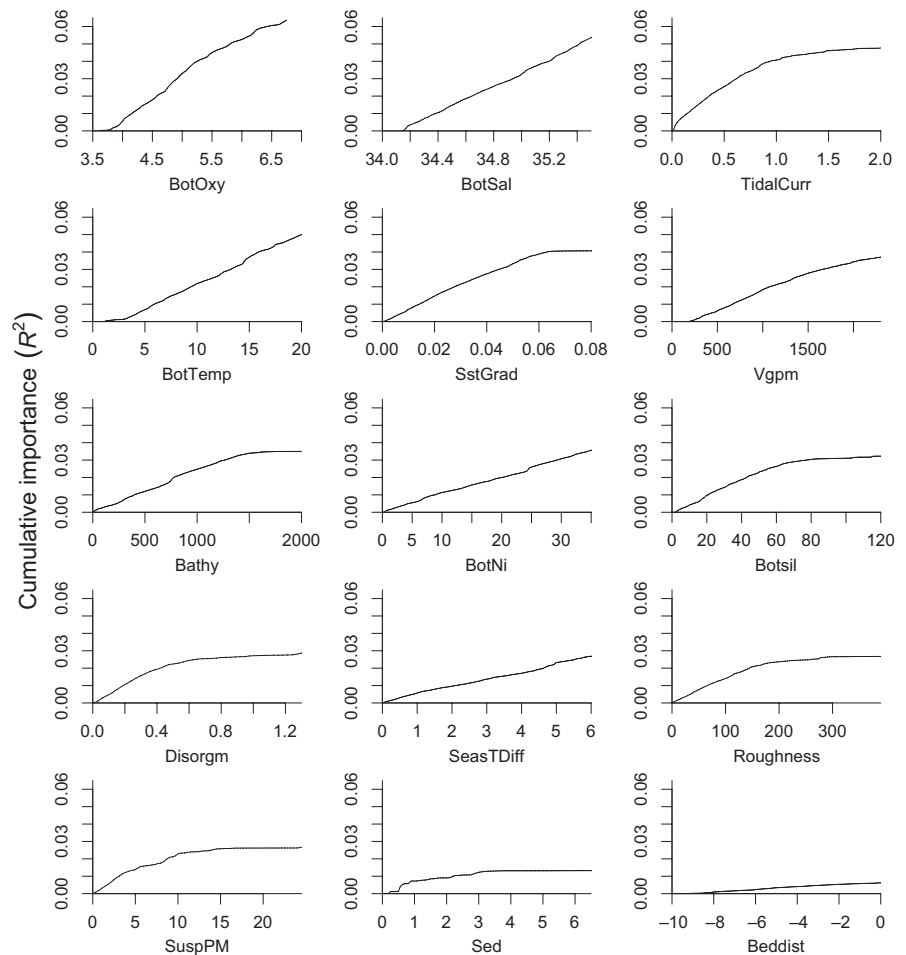

FIGURE 5 Functions fitted by the final GF model using 13,917 observations and indicating relative species turnover along the range of each predictor
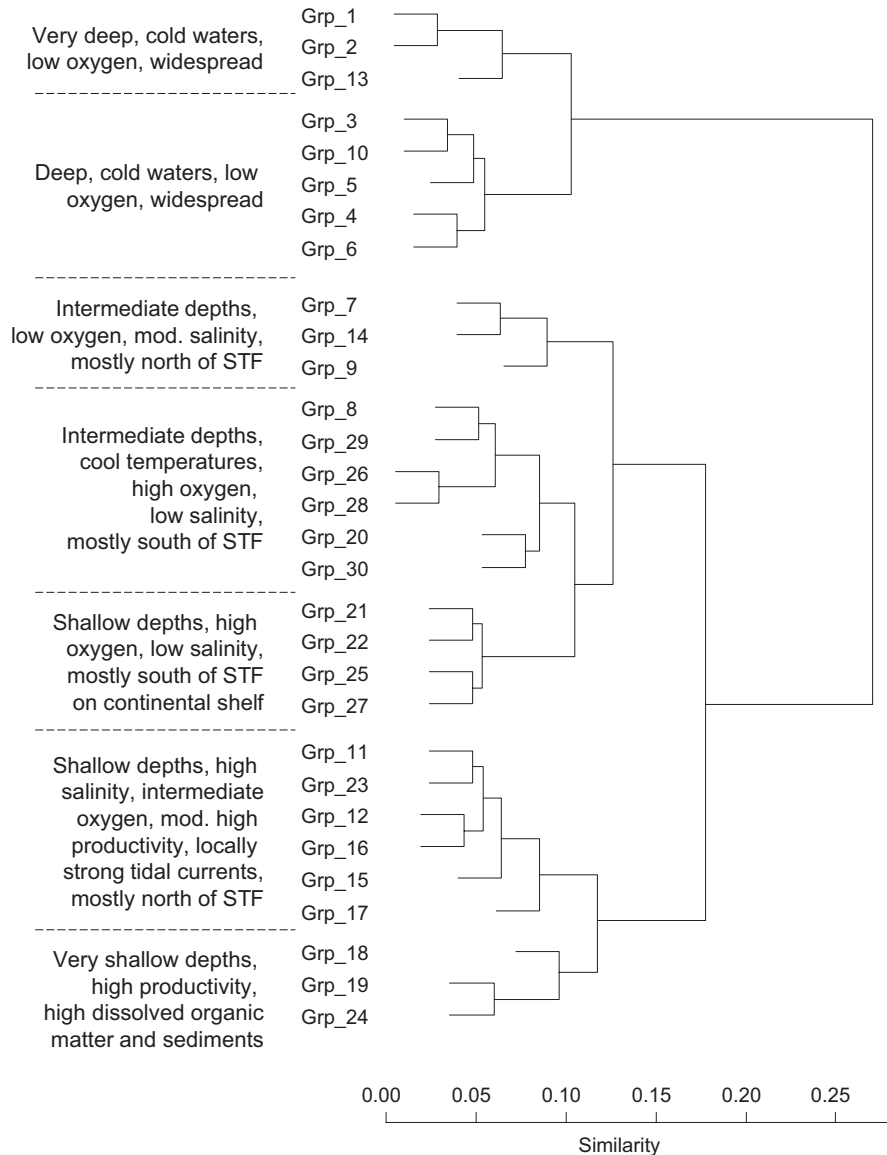
FIGURE 6 Dendrogram describing similarities between 30 demersal fish groups defined by classification of transformed environmental layers for the seas within the New Zealand Extended Continental Shelf to a depth of 2,000 m after transformation using a Gradient Forest model fitted to presence–absence data from 13,917 research trawls. See text for a broad description of groups is provided

3.4 | Model validation

Comparison of distances for the final classification of the transformed environmental layers, with an equivalent set of biological distances calculated using biological samples from our training dataset, grouped after allocation to their corresponding classification group, indicates that the classification is highly effective at capturing biological turnover and therefore compositional differences between the underlying biological samples. When calculated using those classification groups represented by 200 or more biological samples, distances between pairs of groups in the transformed environmental space have a correlation of 0.934 with equivalent biological distances between the same pairs of groups (Figure 8a). Similar results were obtained when this comparison was repeated using the 13,917 trawl observations that were withheld from the final model fitting to allow independent evaluation of model performance (correlation = 0.919, Figure 8b). The respective correlations maintained

a high level (0.926 and 0.912, respectively) when this comparison was repeated using all pairwise comparisons between groups supported by 100 or more trawl observations.

4 | DISCUSSION

Results from our analysis indicate that Gradient Forest modelling is capable of effectively combining irregular species distribution data with spatially continuous environmental data layers to create a comprehensive description of spatial variation in species composition and turnover, in our case for demersal fish species in New Zealand's Extended Continental Shelf to depths of 2,000 m. In particular, the environmental transforms generated by our final model allowed the creation of a multiscale spatial classification of marine environments that, when assessed using both the training samples and a large set of independent samples, proved to be highly effective in
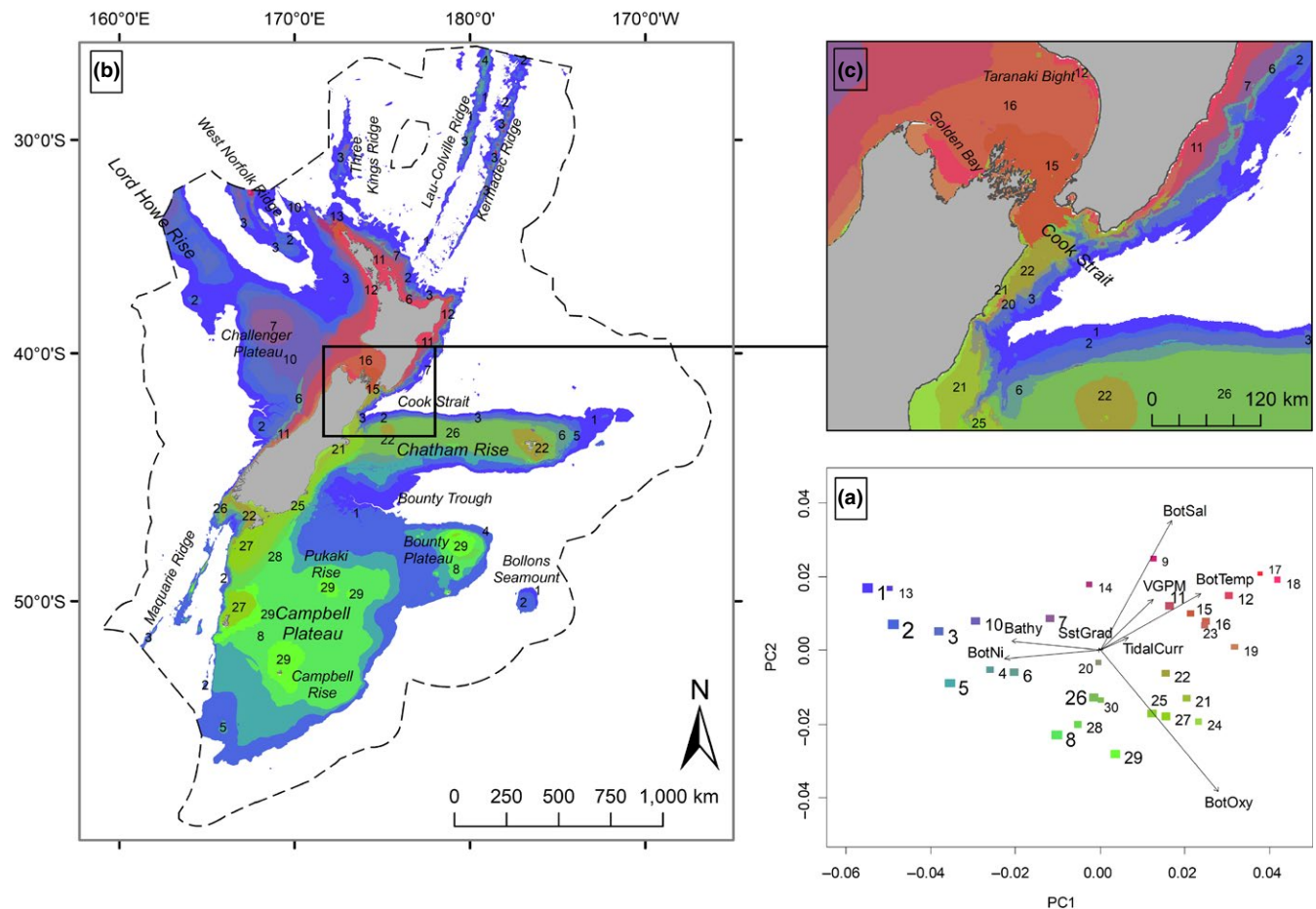
FIGURE 7 Distributions in PCA and geographic space of 30 groups defined by classification of transformed environmental layers for the seas within the New Zealand Extended Continental Shelf (dashed line) to a depth of 2,000 m; transformations were derived from a Gradient Forest model fitted to 13,917 research trawls (Supporting information Figure S2). Colours are based on the first three axes of a PCA analysis applied to the group means for each of the transformed predictor variables, so that similarities/differences in colour correspond broadly to intergroup similarities/differences with respect to the transformed environmental variables. (a) Distributions of groups in PCA space, with vectors indicating correlations with the eight most important environmental predictors and symbol/font size indicating the relative size of the group area; (b) geographic distributions of groups across New Zealand's Extended Continental Shelf (dashed line); (c) geographic distribution of groups at finer scales, centred on Cook Strait [Colour figure can be viewed at wileyonlinelibrary.com]

summarizing both variation in fish assemblage composition and species turnover, as indicated by the high correlation between environmental and compositional distances between pairs of classification groups. In broad terms, this result is consistent with other analyses using Gradient Forests in marine environments (Compton et al. 2013; Thomson et al. 2014), although we note that most of these were carried out over more limited geographic extents and in shallower waters.

Spatial patterns of turnover in New Zealand's demersal fish species described by our 30-group classification were strongly related to latitudinal changes in water mass and their associated frontal features (STF, TF, SAF, Figure 1b; Foster, Givens, Dornan, Dunstan, & Darnell, 2013; Francis et al., 2002; Hill, Lucieer, Barrett, Anderson, & Williams, 2014; Hill et al., 2017). Although environmental differences between classification groups were relatively muted in the deepest waters, differences in temperature, oxygen and salinity become increasingly important in intermediate and shallow

depths. Variation in productivity, sea surface temperature gradients, tidal currents, suspended sediments and dissolved organic matter were important differentiating factors in shallow, inshore waters, where latitudinal differences in environment are in turn reflected in marked differences in fish species assemblages. These results are broadly consistent with those from other New Zealand studies at various spatial scales ranging from regional to national, for example, in Francis et al. (2002), Bull et al. (2001), Beentjes et al. (2002), Kendrick and Francis (2002), Snelder et al. (2007). Using correspondence and cluster analysis, these studies identified depth, and to a lesser extent, water temperature, latitude (as a proxy for temperature and water mass) and major oceanographic features as important variables structuring variation in species composition. While superficially similar to the earlier classification of the oceans around New Zealand by Snelder et al. (2007), our GF classification incorporates important advances through use of a more comprehensive set of ocean climate predictors and an evidence-based approach to
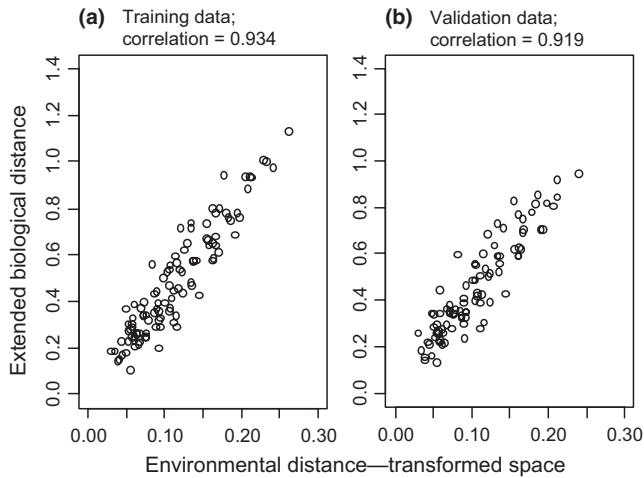
(a) Training data;
correlation = 0.934

(b) Validation data;
correlation = 0.919

Environmental distance—transformed space

FIGURE 8 Extended biological distances of samples grouped according to Gradient Forest model classification with more than 200 observations against distance in transformed environmental space of samples (a) training data (used in the full Gradient Forest model) and (b) independent evaluation data

transformation of predictors prior to classification. This results in a classification that more evenly divides parallel variations in species composition, making it more strongly suited for systematic conservation planning purposes.

Although the species distribution data we used mostly provided adequate spatial coverage of our study area, several large, outlying sections had few or no trawl stations, notably the submarine ridges and steep slopes around the margins of the Campbell and Bounty plateaus. Predictions of fish richness for these areas, produced during a previous study using a subset of the data used here, had wider confidence intervals than more intensively sampled areas (Leathwick et al., 2006), indicating that further limited sampling would most likely improve confidence in our results in these locations. A measure of uncertainty of species turnover could be calculated (similarly to Thomson et al., 2014) although this does not readily expand to assessing confidence in assemblage classification. Alternatively, other modelling methods such as Regions of Common Profile (RCP) could be used to estimate confidence of assemblage classification and which can be validated spatially using the probabilities of occurrence of the individual species of the training samples (Foster et al., 2013; Hill et al., 2017).

4.1 | The effect of variation in numbers of training observations on Gradient Forest models

As a tool, GF appears well suited to the analysis of the often sparse or patchily distributed data that are typically available in marine environments. Although a GF model could accommodate all the ca. 27,800 observations available to us, use of randomly selected independent data subsets indicated that both the number of species effectively modelled and their mean predictive performance (R^2_f) stabilized when using ~25% of the total available dataset (ca. 7,000 observations). Although the mean species predictive performance did not vary greatly with sample number, small but consistent

increases occurred for individual species R^2_f as the size of the training dataset was increased. We interpret this result as most likely indicating that species that have good observational support in the models fitted to small datasets, show further predictive refinement as more observations are added; however, mean predictive performance across all species models fitted when using larger datasets showed little positive trend because improvements in well-fitted species were offset by the lower predictive performance for rarer species that came into play as the size of the training dataset increased.

A common set of environmental predictors (*Bathy*, *BotOxy*, *TidalCurr*, *BotTemp* and *BotSal*) played a dominant role in models fitted with widely varying numbers of observations, although with some resorting of their order of importance with increasing sample size (Bosch, Tyberghein, Deneudt, Hernandez, & de Clerck, 2017; Pitcher et al., 2012). Those predictors showing the largest changes in importance with increasing sample size were predictors that were largely invariant except in a few specific locations, for example, dissolved organic matter (*Disorgm*) which varies only in coastal waters, and sea surface temperature gradient (*SstGrad*) which shows locally high values in areas of ocean mixing and along sharply defined current boundaries, but often with little variation elsewhere. With increasing sample number, these locally divergent sites were better represented in the model, and their ability to influence model outcomes was increased. Therefore, we recommend an explicit examination of the relative influence of predictor variables across a range of sample sizes, as an additional consideration when determining the most appropriate split of data into training and validation subsets to ensure representation of the full environmental gradient of the study area. Here, ca. 7,000 observations were adequate to predict demersal fish turnover across a 2,461,926 km^2 study area although this is likely to differ for other geographic locations and taxa.

An additional consideration when analysing datasets accumulated for purposes other than biodiversity description is GF's more relaxed assumptions about species absence compared to GDM. Whereas the latter was primarily designed to work with data listing all species present at each of some set of sample sites (Ferrier et al., 2007), the individual classification tree regressions fitted in a GF model make the less stringent assumption that nonpositive occurrences for each species can be treated as an assumed absence (Pitcher et al., 2011). While this is less ideal than the use of true presence–absence data, models fitted under this assumption can still produce reasonably robust predictions of species distributions, particularly when the assumed absences are drawn from a broader set of samples for the same biotic group (Elith & Leathwick, 2009; Elith et al., 2006). In the GF models presented here, true presence/absences were used.

4.2 | Critical appraisal of the "final" Gradient Forest model

Although our repeated analyses with subsamples of the training dataset indicated minimal changes in model stability with expansion

of the data beyond ca. 7,000 samples, we adopted a relatively conservative stance, fitting our "final" model with 13,917 observations or half of the available data; the other half we used solely for independent evaluation of our "final" model. We could have fitted models to a higher proportion of the total number of observations (a typical ratio is 75:25), but we had the luxury of a very large dataset; consideration should be given to using a higher fraction of the sample data when using smaller datasets. The 253 individual Random Forest models in this final GF model had relatively high predictive performance values (mean R^2_f of 0.521 and range of 0.3–0.913), and all showed at least some predictive ability. These results are unusual for marine species distribution studies; in a model comparison study, Elith et al. (2006) found that only half of the marine species included in the study had useable models and few had high predictive power even for tree ensemble models (BRT and MAXENT), which consistently outperformed more established regression methods (e.g., GLM, GAM and MARS).

There was a weak curvilinear relationship between numbers of positive observations for species and their predictive performance, that is, more frequently occurring species had better fitting models, although those with the lowest frequencies still had R^2_f values >0.3. Interestingly, there was a broad range of R^2_f (0.41–0.86) for species with moderate–high observations (1,000–4,000), reflecting perhaps the cosmopolitan distribution of some species (e.g., *Bassanago* spp., low R^2_f) and the more aggregated nature of others (e.g., orange roughy, high R^2_f; see Supporting information Table S1). Evidence from other studies indicates that species with limited geographic range and environmental tolerances are generally better modelled than those with greater ranges (Guisan et al., 2013; Morán-Ordóñez et al., 2017; Thomson et al., 2014) because widespread species are less likely to have sharp easily identifiable environmental thresholds that clearly delineate their most suitable environmental conditions (Morán-Ordóñez et al., 2017). Reduced model fit could be influenced by historical events, human activities, population and species dynamics (e.g., recruitment, competition, predation and facilitation) (Elith & Leathwick, 2009) and temporal environmental patterns (e.g., diurnal, tidal and seasonal cycles and fluctuating weather patterns) not accounted for here (Compton et al., 2013; Pitcher et al., 2012). The lack of consideration of these factors in a quantitative manner does not invalidate the use of the GF model for management, but it should be noted that the representation of assemblages shown here is a (spatially and temporally) smoothed representation of the raw data and further work on integrating more explicit predictors into GF modelling would be of interest (Compton et al., 2013; Thomson et al., 2014).

In general, there was a linear relationship between species turnover along the individual environmental gradients, although there was a plateauing in turnover in parts of the range of some variables, for example, the decrease in species turnover at depths >1,500 m, with species composition showing minimal further change with further increases in depth. A lack of steep slopes in the turnover functions indicates that species distributions generally overlap to a high degree, which is consistent with findings from other studies of shelf fish assemblages both in New Zealand and elsewhere (Beentjes et al., 2002; Bianchi, 1991; Farina, Freire, & González-Gurriarán, 1997; Fujita, Inada, & Ishito, 1995).

4.3 | Usefulness of Gradient Forest models for conservation planning

In our view, the strong discrimination of fish distribution patterns across New Zealand's Extended Continental Shelf shown by our classification of the transformed environmental layers produced by our "final" GF analysis makes it well suited as a primary input to systematic spatial conservation planning analyses (e.g., using spatial planning software such as Zonation (Lehtomäki & Moilanen, 2013) or Marxan (Ball, Possingham, & Watts, 2009)). In addition, significant gains can be expected if such analyses are implemented using prioritization tools that, rather than treating all groups as equally different from each other, take account of their varying intergroup similarities, allowing greater priority to be given to more distinctive species assemblages as can be implemented, for example, in Zonation (Leathwick, Moilanen, Ferrier, and Julian 2010).

Given the hierarchical nature of the classification, consideration will be required as to what constitutes the most appropriate level of classification detail for conservation planning purposes. Here, we have chosen to describe the classification at a 30-group level to facilitate communication; however, our testing of correlations between environmental and biological distances at higher levels of classification detail (up to 100 groups) indicate that these levels can provide even greater discrimination of compositional differences and species turnover than those presented here. In addition, using a higher number of classification groups is likely to be more appropriate for a regional scale analysis, particularly for inshore areas where there is a greater heterogeneity in environmental conditions. Alternatively, regional patterns may be better described by analyses using regional subsets of the data, particularly if the mix of factors controlling species turnover varies region by region. Results presented here would be of use for a gap analysis of assemblages currently protected in MPAs (Leaper et al., 2011) and would allow evidence-based targeting of underrepresented assemblages for further sampling and/or protection (Ferrier et al., 2007; Pitcher et al., 2007).

Finally, consideration will be required of which other taxonomic groups should be included in any operational conservation planning; for example, models describing the distributions of macroalgae and benthic invertebrates could be used in conjunction with our model for demersal fish (Thomson et al., 2014). Data for these different taxonomic groups could be analysed using a prioritization analysis with separate spatial layers describing each classification of interest (Geange et al., 2017). Alternatively, GF modelling allows the combination of results from different datasets into a single unified classification through the averaging of the species turnover functions across taxonomic groups (see example in Pitcher et al. 2012), an approach that GF facilitates through its ability to handle

differences in survey methods, sampling devices and/or measurement scales (counts, weights, abundance and presence/absences; Ellis et al., 2012). However, because this type of analysis has not been undertaken to our knowledge using such widely differing taxonomic groups, we are uncertain as to whether this would provide an adequate spatial representation of all the species of interest (C. R. Pitcher, pers. comm. 2018) and suggest this as a topic for future investigation.

5 | CONCLUSION

Gradient Forest modelling provided an effective way to combine research trawl records with a set of relevant environmental data layers to create a generalized model of demersal fish compositional turnover across a very large spatial extent (New Zealand's Extended Continental Shelf to depths of 2,000 m). However, the recent development of Gradient Forest modelling means that some aspects of using its outputs require further investigation. Of note is a lack in understanding of changes in environmental predictor importance with varying scale (geographic and richness of sampling) and best practice for use in conservation planning, although further work is currently underway to further inform these areas. Although an extensive set of sampling locations was available for the training and validation of the model, results here suggest that model performance stabilized when using a relatively low number of samples (~5,000–7,000). This provides a key advantage when modelling marine species where sampling is often irregularly distributed, opportunistic (presence only) and expensive. The resulting spatial description of species composition and turnover in New Zealand demersal fish assemblages will be highly valuable for underpinning evidence-based conservation planning.

DATA ACCESSIBILITY

Species turnover and 30-group classification generated for this study will be made available as raster grids from the NIWA database upon publication.

ORCID

Fabrice Stephenson http://orcid.org/0000-0002-9500-5204

REFERENCES

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Tracey, D. M., Mackay, K. A., & Clark, M. R. (2016). 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. https://doi.org/10.1016/j.dsr.2016.07.006

Arponen, A., Moilanen, A., & Ferrier, S. (2008). A successful community-level strategy for conservation prioritization. *Journal of Applied Ecology*, *45*, 1436–1445. https://doi.org/10.1111/j.1365-2664.2008.01513.x

Bailey, K. D. (1983). Sociological classification and cluster analysis. *Quality and Quantity*, *17*, 251–268.

Ball, I. R., Possingham, H. P., & Watts, M. (2009). *Marxan and relatives: Software for spatial conservation prioritisation. Spatial conservation prioritisation: Quantitative methods and computational tools* (pp. 185–195). Oxford, UK: Oxford University Press.

Beentjes, M. P., Bull, B., Hurst, R. J., & Bagley, N. W. (2002). Demersal fish assemblages along the continental shelf and upper slope of the east coast of the South Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *36*, 197–223. https://doi.org/10.1080/00288330.2002.9517080

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, *42*, 1–20. https://doi.org/10.4319/lo.1997.42.1.0001

Belbin, L. (1993). Environmental representativeness: Regional partitioning and reserve selection. *Biological Conservation*, *66*, 223–230. https://doi.org/10.1016/0006-3207(93)90007-N

Belbin, L., Faith, D. P., & Milligan, G. W. (1992). A comparison of two approaches to beta-flexible clustering. *Multivariate Behavioral Research*, *27*, 417–433. https://doi.org/10.1207/s15327906mbr2703_6

Bianchi, G. (1991). Demersal assemblages of the continental shelf and slope edge between the Gulf of Tehuantepec (Mexico) and the Gulf of Papagayo (Costa Rica). *Marine Ecology Progress Series*, *73*, 121–140. https://doi.org/10.3354/meps073121

Bosch, S., Tyberghein, L., Deneudt, K., Hernandez, F., & de Clerck, O. (2017). In search of relevant predictors for marine species distribution modelling using the MarineSPEED benchmark dataset. *Diversity and Distributions*, *24*, 144–157.

Bradford-Grieve, J. M., Lewis, K. B., & Stanton, B. R. (1991). Advances in New Zealand oceanography, 1967–1991. *New Zealand Journal of Marine and Freshwater Research*, *25*, 429–441. https://doi.org/10.1080/00288330.1991.9516495

Bradford-Grieve, J., Probert, K., Lewis, K., Sutton, P., Zeldis, J., & Orpin, A. (2006). *New Zealand shelf region*. Cambridge, MA: Harvard University Press.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Bull, B., Livingston, M., Hurst, R., & Bagley, N. (2001). Upper-slope fish communities on the Chatham Rise, New Zealand, 1992–99. *New Zealand Journal of Marine and Freshwater Research*, *35*, 795–815. https://doi.org/10.1080/00288330.2001.9517043

Bush, A., Harwood, T., Hoskins, A. J., Mokany, K., & Ferrier, S. (2016). Current uses of beta-diversity in biodiversity conservation: A response to. *Trends in Ecology & Evolution*, *31*, 337. https://doi.org/10.1016/j.tree.2016.02.020

CANZ. (2008). *New Zealand Regional Bathymetry*, NIWA Chart.

CARS2009. (2009). *CSIRO Atlas of Regional Seas* [Online]. CSIRO. Retrieved from: www.cmar.csiro.au/cars (Accessed 14 December 2017).

Carter, L. 2001. *Currents of change: the ocean flow in a changing world* [Online]. Wellington, New Zealand: NIWA. Retrieved from: https://www.niwa.co.nz/publications/wa/vol9-no4-December-2001/currents-of-change-the-ocean-flow-in-a-changing-world (Accessed 05 February 2018).

Compton, T. J., Bowden, D. A., Pitcher, C. R., Hewitt, J. E., & Ellis, N. (2013). Biophysical patterns in benthic assemblage composition across contrasting continental margins off New

Zealand. *Journal of Biogeography*, *40*, 75–89. https://doi.org/10.1111/j.1365-2699.2012.02761.x

Dunstan, P. K., Althaus, F., Williams, A., & Bax, N. J. (2012). Characterising and predicting benthic biodiversity for conservation planning in deepwater environments. *PLoS ONE*, *7*, e36558. https://doi.org/10.1371/journal.pone.0036558

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., … Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*, 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, *93*, 156–168. https://doi.org/10.1890/11-0252.1

Farina, A., Freire, J., & González-Gurriarán, E. (1997). Demersal fish assemblages in the Galician continental shelf and upper slope (NW Spain): Spatial structure and long-term changes. *Estuarine, Coastal and Shelf Science*, *44*, 435–454. https://doi.org/10.1006/ecss.1996.0148

Ferrier, S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology*, *51*, 331–363. https://doi.org/10.1080/10635150252899806

Ferrier, S., Manion, G., Elith, J., & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, *13*, 252–264. https://doi.org/10.1111/j.1472-4642.2007.00341.x

Ferrier, S., Powell, G. V., Richardson, K. S., Manion, G., Overton, J. M., Allnutt, T. F., … Faith, D. P. (2004). Mapping more of terrestrial biodiversity for global conservation assessment. *BioScience*, *54*, 1101–1109. https://doi.org/10.1641/0006-3568(2004)054[1101:MMOTBF]2.0.CO;2

Foster, S., Givens, G., Dornan, G., Dunstan, P., & Darnell, R. (2013). Modelling biological regions from multi-species and environmental data. *Environmetrics*, *24*, 489–499. https://doi.org/10.1002/env.2245

Francis, M. P., Hurst, R. J., McArdle, B. H., Bagley, N. W., & Anderson, O. F. (2002). New Zealand demersal fish assemblages. *Environmental Biology of Fishes*, *65*, 215–234. https://doi.org/10.1023/A:1020046713411

Fujita, T., Inada, T., & Ishito, Y. (1995). Depth-gradient structure of the demersal fish community on the continental shelf and upper slope off Sendai Bay, Japan. *Marine Ecology Progress Series*, *118*, 13–23. https://doi.org/10.3354/meps118013

Geange, S. W., Leathwick, J., Linwood, M., Curtis, H., Duffy, C., Funnell, G., & Cooper, S. (2017). Integrating conservation and economic objectives in MPA network planning: A case study from New Zealand. *Biological Conservation*, *210*, 136–144. https://doi.org/10.1016/j.biocon.2017.04.011

Gorman, R. M., Bryan, K. R., & Laing, A. K. (2003). Wave hindcast for the New Zealand region: Deep-water wave climate. *New Zealand Journal of Marine and Freshwater Research*, *37*, 589–612. https://doi.org/10.1080/00288330.2003.9517191

Goslee, S. C., & Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, *22*, 1–19.

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*, 1424–1435. https://doi.org/10.1111/ele.12189

Harrison, S., Ross, S. J., & Lawton, J. H. (1992). Beta diversity on geographic gradients in Britain. *Journal of Animal Ecology*, *61*, 151–158. https://doi.org/10.2307/5518

Hawkins, C. P., Norris, R. H., Gerritsen, J., Hughes, R. M., Jackson, S. K., Johnson, R. K., & Stevenson, R. J. (2000). Evaluation of the use of landscape classifications for the prediction of freshwater biota: Synthesis and recommendations. *Journal of the North American Benthological Society*, *19*, 541–556. https://doi.org/10.2307/1468113

Hewitt, J. E., Wang, D., Francis, M., Lundquist, C., & Duffy, C. (2015). Evaluating demersal fish richness as a surrogate for epibenthic richness in management and conservation. *Diversity and Distributions*, *21*, 901–912. https://doi.org/10.1111/ddi.12336

Hijmans, R. J., & van Etten, J. 2012. raster: Geographic analysis and modeling with raster data. R package version 2.0-12. Retrieved from: http://CRAN.R-project.org/package=raster

Hill, N. A., Foster, S. D., Duhamel, G., Welsford, D., Koubbi, P., & Johnson, C. R. (2017). Model-based mapping of assemblages for ecology and conservation management: A case study of demersal fish on the Kerguelen Plateau. *Diversity and Distributions*, *23*, 1216–1230. https://doi.org/10.1111/ddi.12613

Hill, N. A., Lucieer, V., Barrett, N. S., Anderson, T. J., & Williams, S. B. (2014). Filling the gaps: Predicting the distribution of temperate reef biota using high resolution biological and acoustic data. *Estuarine Coastal and Shelf Science*, *147*, 137–147. https://doi.org/10.1016/j.ecss.2014.05.019

Jenkins, C. J. (2010). Seafloor substrates. INSTAAR, University of Colorado, Boulder, CO. Retrieved from http://csdms.colorado.edu/wiki/Data:DBSEABED

Kendrick, T. H., & Francis, M. P. (2002). Fish assemblages in the Hauraki Gulf, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *36*, 699–717. https://doi.org/10.1080/00288330.2002.9517124

Leaper, R., Hill, N. A., Edgar, G. J., Ellis, N., Lawrence, E., Pitcher, C. R., … Thomson, R. (2011). Predictions of beta diversity for reef macroalgae across southeastern Australia. *Ecosphere*, *2*, 1–18.

Leathwick, J., Elith, J., Francis, M., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: An analysis using boosted regression trees. *Marine Ecology Progress Series*, *321*, 267–281. https://doi.org/10.3354/meps321267

Leathwick, J., Moilanen, A., Ferrier, S., & Julian, K. (2010). Complementarity-based conservation prioritization using a community classification, and its application to riverine ecosystems. *Biological Conservation*, *143*, 984–991. https://doi.org/10.1016/j.biocon.2010.01.012

Leathwick, J., Rowden, A., Nodder, S., Gorman, R., Bardsley, S., Pinkerton, M., … Goh, A. (2012). *A Benthic-optimised marine environment classification (BOMEC) for New Zealand waters. New Zealand Aquatic Environment and Biodiversity Report; ISSN 1176-9440.*Wellington, New Zealand: Ministry of Fisheries.

Leathwick, J., Snelder, T., Chadderton, W., Elith, J., Julian, K., & Ferrier, S. (2011). Use of generalised dissimilarity modelling to improve the biological discrimination of river and stream classifications. *Freshwater Biology*, *56*, 21–38. https://doi.org/10.1111/j.1365-2427.2010.02414.x

Legendre, P., Borcard, D., & Peres-Neto, P. R. (2005). Analyzing beta diversity: Partitioning the spatial variation of community composition data. *Ecological Monographs*, *75*, 435–450. https://doi.org/10.1890/05-0549

Lehtomäki, J., & Moilanen, A. (2013). Methods and workflow for spatial conservation prioritization using Zonation. *Environmental Modelling & Software*, *47*, 128–137. https://doi.org/10.1016/j.envsoft.2013.05.001

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*, 18–22.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2017). cluster: Cluster analysis basics and extensions. *R Package Version*, *2*, 6.

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, *405*, 243. https://doi.org/10.1038/35012251

McKnight, M. W., White, P. S., McDonald, R. I., Lamoreux, J. F., Sechrest, W., Ridgely, R. S., & Stuart, S. N. (2007). Putting beta-diversity on the map: Broad-scale congruence and coincidence in

the extremes. *PLoS Biology*, *5*, e272. https://doi.org/10.1371/journal.pbio.0050272

Mitchell, J. S., Mackay, K. A., Neil, H. L., Mackay, E. J., Pallentin, A., & Notman, P. (2012). *Undersea New Zealand, 1:5,000,000. NIWA Chart, Miscellaneous Series No. 92.* .

Morán-Ordóñez, A., Lahoz-Monfort, J. J., Elith, J., & Wintle, B. A. (2017). Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: What factors influence the reliability of predictions? *Global Ecology and Biogeography*, *26*, 371–384. https://doi.org/10.1111/geb.12545

Mumby, P. J., Broad, K., Brumbaugh, D. R., Dahlgren, C., Harborne, A. R., Hastings, A., ... Sanchirico, J. N. (2008). Coral reef habitats as surrogates of species, ecological functions, and ecosystem services. *Conservation Biology*, *22*, 941–951. https://doi.org/10.1111/j.1523-1739.2008.00933.x

Nekola, J. C., & White, P. S. (1999). The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, *26*, 867–878. https://doi.org/10.1046/j.1365-2699.1999.00305.x

Nekola, J. C., & White, P. S. (2002). Conservation, the two pillars of ecological explanation, and the paradigm of distance. *Natural Areas Journal*, *22*, 305–310.

Omernik, J. M., & Bailey, R. G. (1997). Distinguishing between watersheds and ecoregions. *JAWRA Journal of the American Water Resources Association*, *33*, 935–949. https://doi.org/10.1111/j.1752-1688.1997.tb04115.x

Pinkerton, M. (2016). *Ocean colour satellite observations of phytoplankton in the New Zealand EEZ, 1997–2016. Prepared for the Ministry for the Environment.* Wellington, New Zealand: NIWA.

Pinkerton, M. H., Moore, G. F., Lavender, S. J., Gall, M. P., Oubelkheir, K., Richardson, K. M., ... Aiken, J. (2006). A method for estimating inherent optical properties of New Zealand continental shelf waters from satellite ocean colour measurements. *New Zealand Journal of Marine and Freshwater Research*, *40*(2), 227–247. https://doi.org/10.1080/00288330.2006.9517417

Pinkerton, M. H., Richardson, K. M., Boyd, P. W., Gall, M. P., Zeldis, J., Oliver, M. D., & Murphy, R. J. (2005). Intercomparison of ocean colour band-ratio algorithms for chlorophyll concentration in the Subtropical Front east of New Zealand. *Remote Sensing of Environment*, *97*, 382–402. https://doi.org/10.1016/j.rse.2005.05.004

Pitcher, C. R., Doherty, P., Arnold, P., Hooper, J., Gribble, N., Bartlett, C., ... Yearsley, G. (2007). Seabed Biodiversity on the Continental Shelf of the Great Barrier Reef World Heritage Area. AIMS/CSIRO/QM/QDPI CRC Reef Research Task Final Report, 320.

Pitcher, C., Ellis, N., & Smith, S. (2011). Example analysis of biodiversity survey data with R package gradientForest.

Pitcher, R. C., Lawton, P., Ellis, N., Smith, S. J., Incze, L. S., Wei, C. L., ... Snelgrove, P. V. (2012). Exploring the role of environmental variables in shaping patterns of seabed biodiversity composition in regional-scale ecosystems. *Journal of Applied Ecology*, *49*, 670–679. https://doi.org/10.1111/j.1365-2664.2012.02148.x

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ridgway, K., Dunn, J., & Wilkin, J. (2002). Ocean interpolation by four-dimensional weighted least squares—Application to the waters around Australasia. *Journal of Atmospheric and Oceanic Technology*, *19*, 1357–1375. https://doi.org/10.1175/1520-0426(2002)019<1357:OIBFDW>2.0.CO;2

Roff, J. C., Taylor, M. E., & Laughren, J. (2003). Geophysical approaches to the classification, delineation and monitoring of marine habitats and their communities. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *13*, 77–90. https://doi.org/10.1002/(ISSN)1099-0755

Shmida, A., & Wilson, M. V. (1985). Biological determinants of species diversity. *Journal of Biogeography*, *12*, 1–20. https://doi.org/10.2307/2845026

Snelder, T. H., & Biggs, B. J. (2002). Multiscale river environment classification for water resources management. *JAWRA Journal of the American Water Resources Association*, *38*, 1225–1239. https://doi.org/10.1111/j.1752-1688.2002.tb04344.x

Snelder, T. H., Leathwick, J. R., Dey, K. L., Rowden, A. A., Weatherhead, M. A., Fenwick, G. D., ... Hadfield, M. G. (2007). Development of an ecologic marine classification in the New Zealand region. *Environmental Management*, *39*, 12–29. https://doi.org/10.1007/s00267-005-0206-2

Socolar, J. B., Gilroy, J. J., Kunin, W. E., & Edwards, D. P. (2016). How should beta-diversity inform biodiversity conservation? *Trends in Ecology & Evolution*, *31*, 67–80. https://doi.org/10.1016/j.tree.2015.11.005

Thomson, R. J., Hill, N. A., Leaper, R., Ellis, N., Pitcher, C. R., Barrett, N. S., & Edgar, J. G. (2014). Congruence in demersal fish, macroinvertebrate, and macroalgal community turnover on shallow temperate reefs. *Ecological Applications*, *24*, 287–299. https://doi.org/10.1890/12-1549.1

Walters, R. A., Goring, D. G., & Bell, R. G. (2001). Ocean tides around New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *35*, 567–579. https://doi.org/10.1080/00288330.2001.9517023

BIOSKETCH

Fabrice Stephenson and **John R Leathwick** are quantitative ecologists with interests in modelling spatial patterns of biodiversity.

Shane W Geange, Ashley A Rowden and **Carolyn Lundquist** are marine ecologists with interests in conservation and marine spatial planning.

Judi E. Hewitt and **Richard H Bulmer** are benthic ecologists with interests in explaining biodiversity patterns at a variety of scales.

Owen F Anderson is a fisheries scientist and spatial modeller with an interest in the effects of fishing on associated species.

Author contributions: C.L., J.R.L. and F.S. conceived the ideas; C.L., O.F.A, J.E.H and A.A.R collated the data; J.R.L. and F.S. analysed the data; F.S. and J.R.L. led the writing; and C.L., O.F.A, J.E.H, R.H.B, S.W.G and A.A.R contributed to the writing.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.