

# Inclusion of biotic variables improves predictions of environmental niche models

Fabrice Stephenson<sup>1,2</sup>  | Rebecca V. Gladstone-Gallagher<sup>3</sup> | Richard H. Bulmer<sup>1</sup> | Simon F. Thrush<sup>3</sup> | Judi E. Hewitt<sup>1,4</sup>

<sup>1</sup>National Institute of Water and Atmosphere (NIWA), Hamilton, New Zealand

<sup>2</sup>School of Science, University of Waikato, Hamilton, New Zealand

<sup>3</sup>Institute of Marine Science, University of Auckland, Auckland, New Zealand

<sup>4</sup>Department of Statistics, University of Auckland, Auckland, New Zealand

## Correspondence

Fabrice Stephenson, School of Science, University of Waikato, Hamilton, New Zealand.

Email: [fabrice.stephenson@waikato.ac.nz](mailto:fabrice.stephenson@waikato.ac.nz)

## Funding information

National Institute of Water and Atmospheric Research Coast and Oceans SSIF, Grant/Award Number: COME2001 and COME2101

Editor: Gwen Iacona

## Abstract

**Aim:** Species Distribution Models (SDMs) are correlative models that predict the occurrence or abundance of species in relation to predictor variables. SDMs have become an important part of resource management and conservation biology yet they rarely incorporate species' biology or demography into their predictions. To explore the possible influence of biotic relationships in explaining patterns of species' distribution, abundance and explanatory power of SDMs, we chose two intertidal shellfish species with overlapping but different environmental preferences (*Austrovenus stutchburyi* and *Macomona liliana*) and modelled their distributions with and without biotic variables.

**Location:** New Zealand.

**Methods:** The relationship between environmental and biotic variables on the abundance of our two species was investigated using Boosted Regression Trees (BRTs) with increasing model complexity: (1) BRT models using environmental variables were fitted to each species; (2) BRT models using environmental variables and the co-occurring abundance of the study taxa not being modelled were fitted; (3) BRT models using environmental variables, the co-occurring abundance and the estimated abundance of the species' patch of the study taxa not being modelled were fitted.

**Results:** A strong, non-linear effect of the abundance of *Austrovenus* on *Macomona* was observed but only a weak effect of *Macomona* on *Austrovenus*. The inclusion of biotic variables improved the model fit metrics for both species, as assessed by withheld evaluation data, markedly so for *Macomona*. The overall deviance explained by the models increased, the correlation of predicted vs observed abundance data increased and the variability in these measures decreased.

**Main conclusions:** The combination of the improvement in model performance and changes in the influence of variables with the inclusion of biotic variables is of importance when predicting into unsampled space (e.g. when predicting impacts of climate change). Our approach improves classic SDMs by integrating ecological theories of how species interactions can alter species distributions across environmental gradients.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

## KEYWORDS

biotic relationships, Boosted Regression Trees, non-linear relationships, species distribution models

## 1 | INTRODUCTION

Characterization of biodiversity patterns across multiple scales (varying from meters to hundreds of kilometres) is required for effective management of ecosystems (Ferrier et al., 2007; Margules & Pressey, 2000). One tool that can provide estimates of biodiversity patterns are species distribution models (SDMs, sometimes also termed Ecological niche models (ENMs). SDMs are correlative models that predict the occurrence of species in relation to environmental variables. By relating species' data to environmental predictor variables, the abundance or probability of taxa occurrence can be estimated along with a characterization of the environmental drivers of species distributions. SDMs have become an important part of resource management and conservation biology. Such models are capable of filling knowledge gaps on spatial and temporal distributions and have been used to predict suitable habitat for a range of species (Elith et al., 2006; Weber et al., 2017). However, SDMs rarely incorporate species' biology or demography into their predictions.

Fundamental niches represent all the environmental conditions where a species can live, whereas the realized niche is where the species actually lives (Hutchinson, 1978). SDMs are models of a species' observed distribution, they therefore represent realized niche (the environments where a species is found at the time of sampling), as opposed to the (potentially broader) fundamental niche (the environments where a species can be found, or where the abiotic environment is appropriate for the survival) (Araújo & Guisan, 2006; Soberón, 2007). Despite SDMs explicitly representing a realized niche, the effect of facilitation, predation, competition or biogenic habitat modification in modifying the realized niche is often overlooked and not included in estimates of SDMs.

Species' distributions are driven by both the biotic and abiotic environment; distributions are dictated by physical environmental gradients, but these limits can be restricted or expanded by both negative and positive species interactions. Facilitation can expand the fundamental niche (so the realized niche is larger than the fundamental), whilst competition and predation can reduce the fundamental niche (so realized niche is smaller than the fundamental) (He & Bertness, 2014; Louthan et al., 2015; Mason et al., 2014; Stachowicz, 2012). Given the role of biotic modifiers in shaping the realized niche, accounting for these interactions (or proxies thereof) in a quantitative way is crucial when predicting into unsampled environmental space (e.g. future predictions with climate change, or predictions to unsampled populations) (Araújo & Luoto, 2007; Dormann et al., 2018).

Despite the potential importance of including biotic interactions in SDMs few examples exist compared to the "usual" environmental only predictions. This may in part be due to the difficulty in measuring biotic interactions in a quantitative way. Here, we assume that

biotic variables (e.g. co-occurring species abundances) can represent biotic interactions noting that this may not always be the case (Blanchet et al., 2020; Poggiato et al., 2021) (see discussion for further detail). There are three main ways SDMs can attempt to account for biotic interaction (but see Dormann et al. (2018) for a comprehensive review of methods and considerations): (1) incorporating a biotic variable (representing a biotic interaction) as a predictor in an environmental model (e.g. as in Leathwick & Austin, 2001); (2) measuring partial correlations of residuals from a model controlling for environmental effects (e.g. as in Harris, 2016); and (3) joint SDMs (jSDMs), which can account for multiple species (assumed biotic) interactions, environmental covariates, species traits and phylogenetic relationships through a hierarchical Bayesian modelling approach (e.g. as in Ovaskainen et al., 2017). The latter method shows considerable promise in the ability to account for biotic interactions across multiple species (i.e. in community data). Furthermore, the method assumes that correlations in species' co-occurrence or abundance patterns are symmetrical (Ovaskainen & Abrego, 2020), that is that there is an equal relationship between two species as opposed to an asymmetrical relationship where one species may affect the other but not vice versa. This assumption may not be adequate for all species' relationships (e.g. where a species is reliant on another facilitator for substrate or where there are competitive interactions among the species). However, the simpler approach of incorporating a biotic variable as a predictor in an environmental model (e.g. as in Leathwick & Austin, 2001) but using flexible machine learning models may be best suited to account for these types of interactions.

As a case study to explore how the addition of biotic variables (which are hypothesized to represent species interactions) improves explanatory power of SDMs, we chose two common estuarine species, the New Zealand cockle *Austrovenus stutchburyi* (hereafter referred to as *Austrovenus*) and the wedge shell *Macomona liliana* (hereafter referred to as *Macomona*). These species modify soft-sediment seafloor habitats, increase biodiversity, alter benthic-pelagic coupling, and deliver valued ecosystem services (Lohrer et al., 2012, 2016; Maire et al., 2010; Rullens et al., 2019; Thrush et al., 2006). *Austrovenus* and *Macomona* show strong distributional changes across environmental gradients and they both have differing but overlapping environmental preferences (Douglas et al., 2019; Ellis et al., 2017; Hewitt et al., 2009; Thrush et al., 2003).

The intra- and inter-species interactions of *Macomona* and *Austrovenus* change over the course of the animals' life cycle. As adults, *Macomona* live at depth (up to 15 cm in the sediment) and mainly deposit feed via a long inhalant siphon at the sediment surface, whilst their juveniles live near the surface sediments. At high enough densities (e.g. 12 adults per 23 cm diameter core, Thrush et al., 1996), the adult's feeding activities at the surface can negatively impact the juvenile conspecifics, though this interaction is

highly context dependent (Hewitt, Thrush, et al., 1997; Thrush et al., 1996, 2000). *Macomona* adults may also disrupt the settlement of juvenile *Austrovenus* when they exist at high enough densities (in the same way they disrupt juvenile conspecifics; Thrush et al., 1996, 2000). As adults and juveniles, *Austrovenus* live and suspension feed at the sediment surface, armouring and stabilizing the bed at high densities. When at densities greater than  $\sim 1200$  adults  $m^{-2}$ , *Austrovenus* essentially occur shell to shell in the top 2 cm of the sediment and this could disrupt the feeding of *Macomona*. Given the autecology of these species, the net outcome of the biotic interactions on total abundance of these species will likely be asymmetrical (i.e. that *Austrovenus* are likely to have a larger effect on *Macomona* abundance than vice versa).

Recent work investigating species interactions using jSDMs between *Austrovenus* and *Macomona* (and other soft-sediment benthic taxa) revealed that there were positive and negative interactions between the two species (Kraan, Thrush, et al., 2020). Here, we explore the influence of (potentially asymmetrical) biotic interactions in explaining patterns of species' distribution and total abundance and explanatory power of flexible machine learning models. An often neglected factor in the analysis of species' distribution patterns is the multi-scale nature of the data (Kraan et al., 2015); biotic interactions are often thought to dominate at small scales whereas environmental influences are often thought to dominate at broader scales (Soberón, 2007). Here, we investigate not only the inclusion of abundance information of co-occurring taxa, but also the role of species

patches (i.e. broad scale aggregations of species) as an important meso-scale explanatory variable. Our approach integrates and improves classic SDMs with ecological theories of how other species may alter species distributions across environmental gradients.

## 2 | METHODS

### 2.1 | Data collection

Macrobenthic fauna and associated environmental variables were sampled during austral summer 2012 in Kaipara, Manukau, and Tauranga harbours, North Island, New Zealand (Kraan, Greenfield, et al., 2020) (Figure 1). Full details on data collection, curation and processing methods are available in Kraan, Greenfield, et al. (2020). Briefly, in each harbour, macrobenthic fauna were collected and abundance recorded from 400 cores (13 cm diameter, 20 cm deep) sampled along four 1000 m transects (spaced 100 m apart) during low tide from the high- to low-water mark (Figure 1). Sampling points along the transects were spaced at distances of 1, 5, 10, 30, 50, 100, 500 and 1000 m (Kraan et al., 2015). Only samples containing at least one of either of our study taxa (*Austrovenus* or *Macomona*, all individuals  $>1$  mm, that is adults and juveniles) were retained for further analysis ( $n = 853$ ).

Key environmental variables that are known to shape our study species' distributions (Hewitt et al., 2005; Thrush et al.,

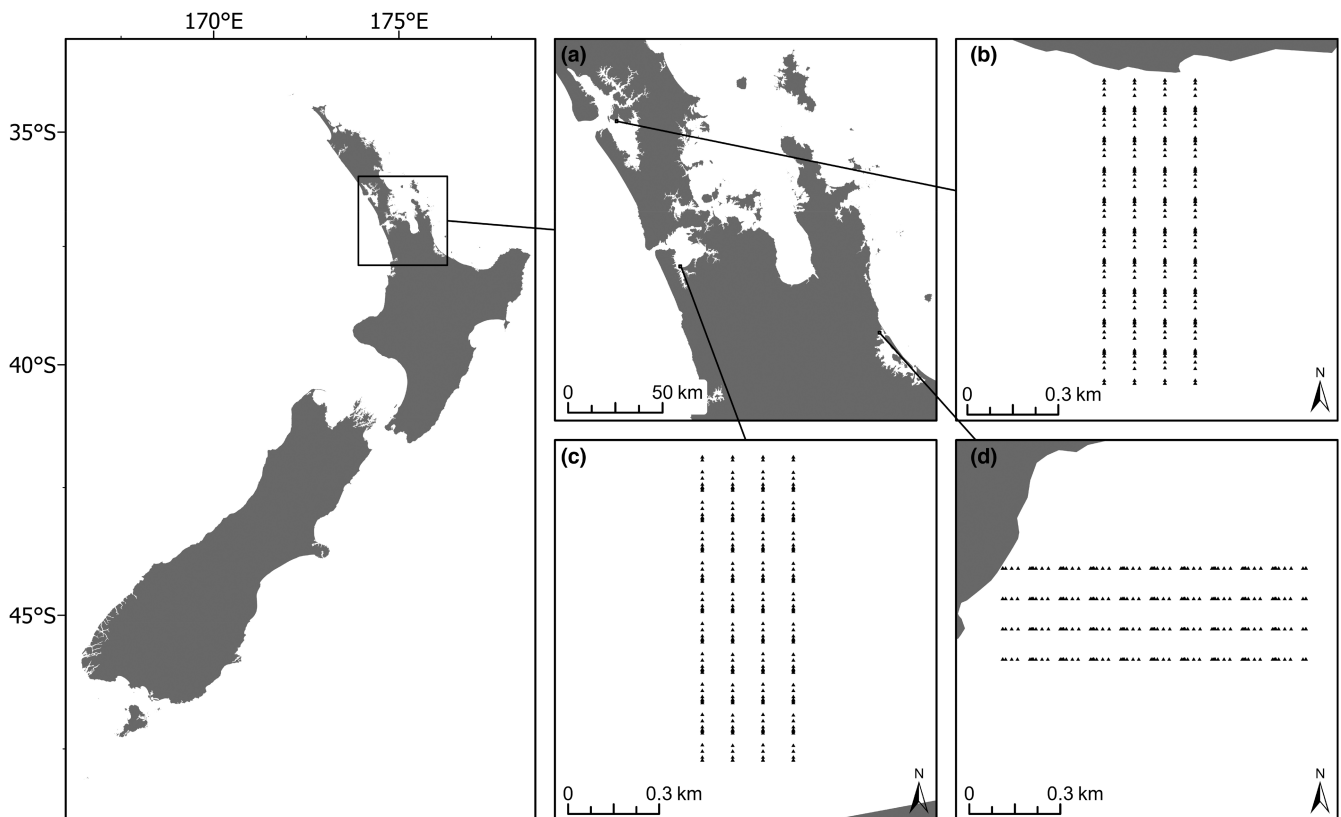


FIGURE 1 Location of macrofaunal sampling for three estuaries in New Zealand (a): Kaipara (b), Manukau (c) and Tauranga (d)

2003) were collected at each sampling point. This included chlorophyll *a* (*Chla*) and pheophytin measurements (*Pheophytin*), sediment grain-size distributions (*Mud*), organic content of the sediment (*LOI*) and visible sandflat parameters (e.g. the coverage of seagrass (*Seagrass*), shellhash (*Shellhash cover*), and bare sand estimated within each photograph of the sample site prior to macrofauna sampling (Table 1). In addition to environmental variables, we explored the influence of biotic variables in explaining patterns of each of our study taxa. This included the co-occurring abundance of the study taxa not being modelled (i.e. *Macomona* or *Austrovenus*), as well as a spatially smoothed representation of this data used to represent a “patch” (i.e. an area of species aggregation, Table 1). While both *Macomona* and *Austrovenus* have been demonstrated to exhibit patchiness on multiple spatial scales (Hewitt, Legendre, et al., 1997), large patches ca. 70 m diameter are documented and likely represent longer term spatial structures generated by ecological processes (e.g. recruitment, habitat facilitation and predation) (Hewitt, Legendre, et al., 1997). Species patches were estimated using Inverse Distance Weighting (IDW) of the co-occurring abundance sample locations with a circular search radius of 70 m diameter.

## 2.2 | Statistical analyses

The relationship between environmental and biotic variables on the abundance of our two species was investigated using Boosted Regression Trees (BRTs) with increasing model complexity. BRT models combine 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).

Firstly, BRT models using environmental variables were fitted to each species (referred throughout as Env models). Secondly, in order to explore the effect of including biotic variables, BRT models using environmental variables and the co-occurring abundance of the study taxa not being modelled were fitted (referred throughout as Env + species models). Finally, BRT models using environmental variables, the co-occurring abundance and the estimated abundance of the “species’ patch” of the study taxa not being modelled were fitted (referred throughout as Env +species + patch models). 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 Gaussian error distribution for log + 1 transformed abundance of *Macomona* and a Poisson error distribution for abundance of *Austrovenus* (error distribution and transformation were selected based on the shape of species’ abundance histograms and to optimize model performance). BRT models were parametrized following recommendations from Elith et al. (2008): with tree complexity of 3 (allowing 3-way interactions of predictor variables), a learning rate of 0.001 (parameters selected so as to fit between

2000 and 3000 trees for each species’ model), a bag fraction of 0.6 (the proportion of data drawn at random, without replacement for each tree) and random 10-fold cross evaluation.

In order to produce a parsimonious model (and avoid the risk of model overfitting) a variable selection process was undertaken. That is, a model with all environmental predictor variables included for each of the study taxa were subjected to a simplification process whereby environmental variables were removed from the models, one at a time, using the “simplify” function (Elith et al., 2006). This simplification process firstly assesses the relative contributions of each variable in terms of deviance explained, with the lowest contributing variables removed from the model, before the model is refitted with the remaining environmental variables (this is somewhat analogous to a stepwise backward selection in linear modelling). The change in deviance explained that resulted from removing the variable was then examined and the process repeated until all variables were sequentially removed. The final models were created by refitting the model with a reduced variable set that balanced the deviance explained with a reasonable number of predictor variables (final variables selected for each study taxa are shown in Tables 2 and 4). All levels of co-linearity between predictor variables were considered acceptable (Pearson correlation <0.8) for tree-based machine learning methods (Dormann et al., 2013; Elith et al., 2010). Spatial differences in the abundance of study taxa not explained by environmental and biotic variables were accounted for with the inclusion of Estuary as factor in the BRT models. Following model tuning, factor Estuary was only retained in the *Macomona* models.

BRT models were assessed using cross-validated measures of model performance (Compton et al., 2012; Elith et al., 2008). Model performance measures included the deviance explained and the Pearson correlation between predicted and observed abundance. The explained deviance provides a measure of the goodness-of-fit between the predicted and raw values (total deviance) (Compton et al., 2012). The relative influence of each environmental variable in the models was the number of times it was selected for splitting, weighted by the squared improvement to the model as a result of each split (using in-bag data) (Friedman & Meulman, 2003). The association between species abundance and the predictor variables was illustrated using partial dependence plots (i.e. predicted response curve of species abundance across the gradient of the variable of interest when all other variables are held at their means).

BRT models were bootstrapped 100 times for each study taxa. That is, a random “training” sample with a sample size equal to the number of abundance records was drawn with replacement, and a model was constructed with the same settings as the original. This process was repeated 100 times, and at each iteration, predictions were made to the “evaluation” data (the remaining data not randomly selected from the abundance database) allowing model fits to be examined both on the training and evaluation data. An ODMAP (Overview, Data, Model, Assessment and Prediction) standardized protocol (Zurell et al., 2020) is provided

**TABLE 1** Statistical distribution of environmental and biotic predictor variables for Kaipara, Manukau, Tauranga harbours, and across all three harbours, summarized as the mean with the 5 and 95% quantiles in brackets

	Kaipara	Manukau	Tauranga	Overall
Environmental variables				
Chla	5.6 (1.9–12.0)	23.5 (14.1–35.4)	11.1 (6.5–18.2)	14.1 (3.0–30.7)
LOI	0.7 (0.5–1.2)	2.4 (1.0–4.7)	1.8 (1.1–2.9)	1.7 (0.6–3.7)
Mud	6.6 (1.9–15.2)	30.9 (10.8–60.22)	22.6 (14.2–34.1)	21.2 (2.8–47.4)
Pheophytin	8.2 (3.3–16.7)	19.1 (7.0–40.6)	6.5 (2.9–12.9)	11.6 (3.4–29.2)
Seagrass	12.4 (0.0–88.0)	4.5 (0.0–32.4)	23.7 (0.0–58.6)	13.6 (0.0–63.8)
Shellhash cover	2.3 (0.0–9.0)	16.3 (0.0–48.4)	3.2 (0.0–11.0)	7.7 (0.0–37)
Biotic variables				
<i>Macomona</i> (log + 1)	1.6 (0.69–2.45)	0.92 (0.0–2.5)	1.5 (0.7–2.1)	1.4 (0.0–2.4)
<i>Macomona</i> Patch (log + 1)	1.6 (0.7–2.4)	0.90 (0.0–2.4)	1.5 (0.7–2.1)	1.4 (0.0–2.4)
<i>Austrovenus</i>	3.0 (0.0–11.3)	11.0 (0.0–40.8)	3.5 (0.0–13.0)	6.1 (0.0–29.0)
<i>Austrovenus</i> Patch	3.0 (0.0–10.8)	11.0 (0.0–40.1)	3.5 (0.0–13.3)	6.1 (0.0–28.9)

in the Appendix (Table A1). ODMAP protocols reflect the main steps involved in building SDMs and are a structured format for documenting and communicating the models, ensuring transparency and reproducibility (Zurell et al., 2020). The R code used in this analysis is available in the Appendix.

The relationship between the environmental and biotic variables can vary by life stage (e.g. juvenile vs adult environmental and biotic interactions may differ). The effects of different life stages of our study taxa on model fits and predicted environmental and biotic relationships were explored by splitting the study taxa into categories classed as adult (>10 mm shell length) and juvenile (<10 mm shell length) for both *Macomona* and *Austrovenus*. However, there were too few samples to run juvenile models only, and adult models displayed near identical relationships to models including all size classes albeit with slightly lower model fits. We therefore analyse all sizes classes together here, acknowledging that juveniles of our study taxa may not be well represented in our analysis.

### 3 | RESULTS

#### 3.1 | Predicting abundance of *Macomona*

The cross-validated estimates of model performance for the *Macomona* Env model were moderate, explaining 18% of the deviance in the (log + 1 transformed) abundance of *Macomona* and having a Pearson's correlation of 0.45 between predicted and observed abundance of withheld evaluation data (Env, Table 2). There was a large increase in cross-validated estimates of model performance when including co-occurring abundance of *Austrovenus* to the initial environment only model (Env + species, 37% as assessed with the evaluation data, Table 2) and a smaller but consistent increase across model fit metrics with the additional inclusion of *Austrovenus* patch (Env + species + patch, 38%

as assessed with the evaluation data, Table 2). With increasing model complexity, model fit metrics between training data and evaluation data were similar (Table 2) providing some evidence that the models were not overly fitted to the training data. The variability in model fitting metrics (standard deviations calculated from the bootstrap replicates) was low, providing evidence that models were consistently performing across bootstrap samples (Table 2). With the inclusion of biotic variables in the models, the relative difference between training and evaluation model fit metrics reduced. For example, the evaluation Pearson's correlation was 77% of the training Pearson's correlation for the Env model, compared to 91% for the Env + species + patch model (Table 2). Further, the variability in Pearson's correlation (i.e. the SD of the mean bootstrapped BRT models) decreased with the inclusion of biotic variables (Table 2).

The influence of environmental variables for predicting the abundance of *Macomona* differed with increasing model complexity (i.e. from Env to Env + species + patch models, Table 3). Despite changes in variable importance, the shape of the relationships between the abundance of *Macomona* and the environmental variables were not markedly different with increasing model complexity (comparison between Env model, Figure A1; Env + species, Figure A2; Env + species + patch, Figure 2).

The most influential environmental variable for predicting the abundance of *Macomona* in the Env models was mud content (mean contribution 27.1%, Table 3), which displayed a strong positive relationship at low mud values (0%–13% mud, Figure A1) followed by a sharp decline which plateaued at approx. 15% mud (Figure A1). The next most important environmental variables were Chlorophyll *a* (Chla) and differences between estuaries (factor estuary) with a mean contribution of 20.2% and 20.3%, respectively (Table 3); other environmental variables had lower overall importance (3.5%–11.5%, Table 3). The relationship between the abundance of *Macomona* and Chlorophyll *a* was broadly inverse to that of mud content (Figure A1). The abundance of *Macomona* was similar between samples

TABLE 2 Cross-validated estimates of model performance for the bootstrapped BRT models fitted with log + 1 transformed abundance estimates of *Macomona* and abundance estimates of *Austrovenus*

Modelled species		Env	Env + species	Env + species + patch
<i>Macomona</i>	Dev (Train)	0.18 ( $\pm 0.02$ )	0.39 ( $\pm 0.02$ )	0.41 ( $\pm 0.03$ )
	Dev (Eval)	0.18 ( $\pm 0.05$ )	0.37 ( $\pm 0.04$ )	0.38 ( $\pm 0.05$ )
	Cor (Train)	0.58 ( $\pm 0.02$ )	0.71 ( $\pm 0.01$ )	0.73 ( $\pm 0.01$ )
	Cor (Eval)	0.45 ( $\pm 0.06$ )	0.65 ( $\pm 0.04$ )	0.66 ( $\pm 0.01$ )
<i>Austrovenus</i>	Dev (Train)	0.41 ( $\pm 0.04$ )	0.41 ( $\pm 0.03$ )	0.43 ( $\pm 0.03$ )
	Dev (Eval)	0.40 ( $\pm 0.08$ )	0.41 ( $\pm 0.06$ )	0.46 ( $\pm 0.06$ )
	Cor (Train)	0.72 ( $\pm 0.03$ )	0.74 ( $\pm 0.03$ )	0.77 ( $\pm 0.02$ )
	Cor (Eval)	0.60 ( $\pm 0.08$ )	0.60 ( $\pm 0.07$ )	0.66 ( $\pm 0.06$ )

Note: Mean ( $\pm$ SD) deviance explained (dev) and correlation (Cor) between predicted and observed abundance of modelled taxa calculated on training (Train) and on withheld evaluation (Eval) data used in BRT models including environmental predictors only (Env), environmental predictors and abundance of co-occurring species not modelled (Env + species) and environmental predictors, abundance of co-occurring species not modelled and abundance of species not modelled from inferred patches.

from Manukau and Tauranga, but lower in samples from Kaipara (factor Estuary, Figure A1). Weak negative relationships between the abundance of *Macomona* and LOI and Seagrass respectively were observed (Figure A1). Weak positive relationships between the abundance of *Macomona* and Phaeophytin and shellhash respectively were observed (Figure A1).

With the inclusion of co-occurring abundance of *Austrovenus* (Env + species), there was a marked change in variable contributions (Table 3). The abundance of *Austrovenus* was the most influential variable (mean 36.7%, Table 3), displaying a very strong negative relationship even at low abundances (1–3) followed by a much weaker negative relationship along the gradient of *Austrovenus* abundance (Figure A2). The importance of mud content declined the most (–15.2% compared to Env model, Table 3) followed by Chlorophyll *a* (Chl*a*) and organic content (LOI) (–9.6% and –5.3%, respectively, Table 3). The importance of other variables did not change greatly (Table 3).

With the inclusion of the *Austrovenus* patch variable (14.1% importance, Env + species + patch, Table 3) the importance of co-occurring abundance of *Austrovenus* declined (–11.0%, Table 3). The importance of other variables did not differ greatly (Table 3). The relationship between the abundance of *Macomona* and *Austrovenus* patch was similar to that observed with co-occurring abundance of *Austrovenus* (Figure 2).

### 3.2 | Predicting abundance of *Austrovenus*

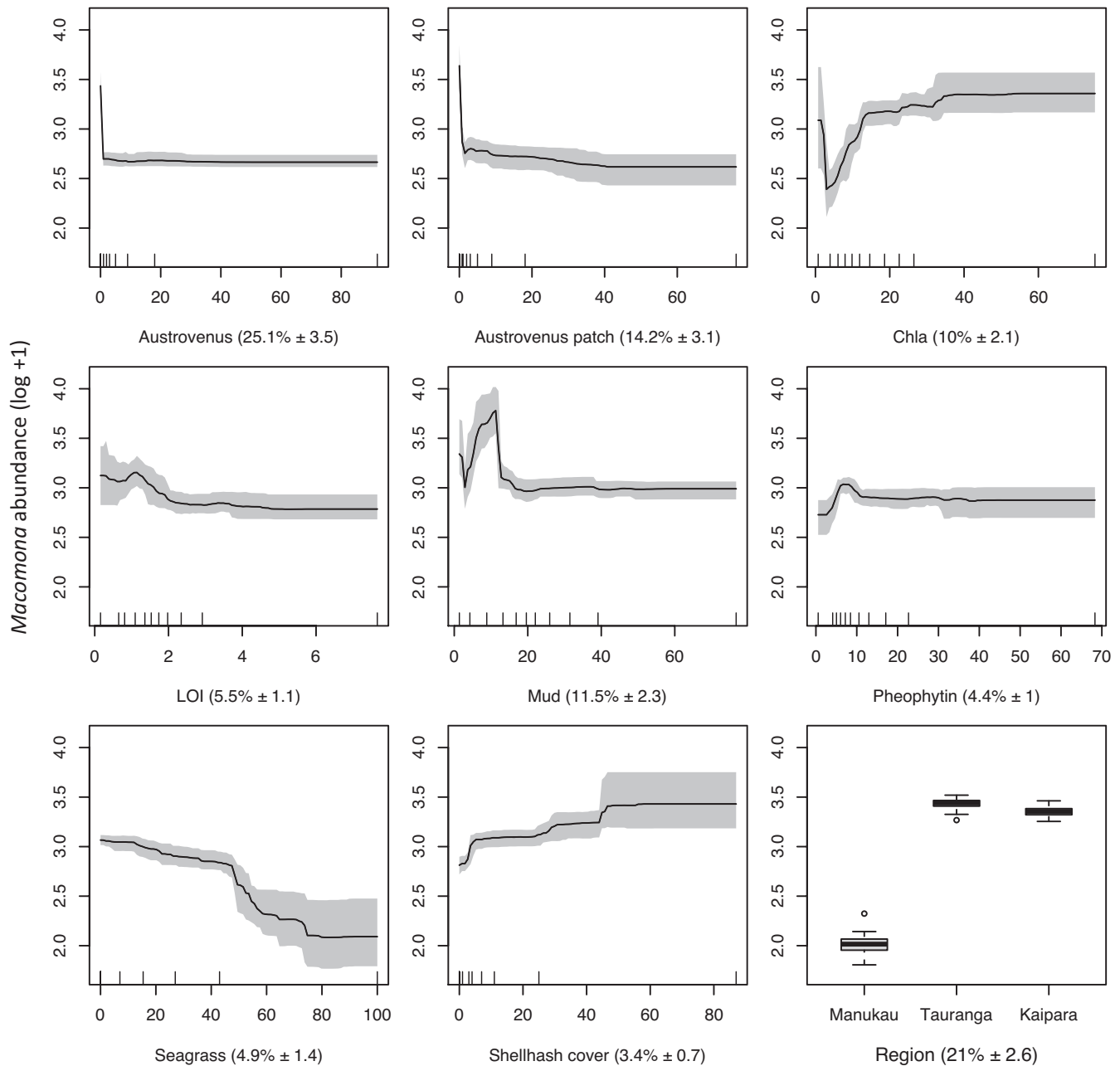
The cross-validated estimates of model performance for the *Austrovenus* environment only model were higher than for the *Macomona* environment only model, explaining 40% of the deviance in the abundance of *Austrovenus* and a Pearson's correlation of 0.60 between predicted and observed abundance of withheld evaluation data (evaluation data, Env, Table 2). However, there was only a modest increase in cross-validated estimates of model performance when including co-occurring abundance of *Macomona* (Env

+ species, 41% as assessed with the evaluation data, Table 2) and *Macomona* patch (Env + species + patch, 46% as assessed with the evaluation data, Table 2). Within models of increasing complexity, model fit metrics between training data and evaluation data were similar (Table 2) providing some evidence that the models were not overly fitted to the training data. In contrast to the *Macomona* models, in the *Austrovenus* models, there were only small (but consistent) decreases in the relative difference between training and evaluation model fit metrics, and the variability around model fit metrics with the inclusion of biotic variables.

The influence of environmental variables for predicting the abundance of *Austrovenus* did not differ as much with increasing model complexity as *Macomona* models did (i.e. from Env to Env + species + patch models, Table 4). As with the *Macomona* models, the shape of the relationships between the abundance of *Austrovenus* and the environmental variables did not differ markedly with increasing model complexity (comparison between Env model, Figure A3; Env + species, Figure A4; Env + species + patch, Figure 3).

The most influential environmental variable for predicting the abundance of *Austrovenus* in the environment only model was shellhash cover (mean contribution 49.6%, Table 4), which displayed a strong positive relationship plateauing at approx. 20% shellhash cover (Figure A3). The next most important environmental variables were mud content with a mean contribution of 24%, followed by Chlorophyll *a* (14.5%) and Seagrass (11.8%, Table 4). The relationship between the abundance of *Austrovenus* and mud content increased sharply at approx. mud values >17% and steadily declined until approx. 50% (Figure A3). A strong positive relationship between abundance of *Austrovenus* and seagrass coverage, and a weaker positive relationship with Chlorophyll *a* was observed (Figure A3).

With the inclusion of co-occurring abundance of *Macomona* (Env +species) there were only small changes in variable contributions (Table 4). The abundance of *Macomona* contributed a modest importance of 5.0% (Table 4), with a weak negative relationship of *Macomona* abundance along the gradient *Austrovenus* abundance (Figure A4). In contrast, the importance of other variables



**FIGURE 2** Partial dependence plots showing the relationships between predictor variables and log + 1 transformed abundance of *Macomona* modelled using bootstrapped BRT (environmental and biotic predictors (Env +species + patch)). Black lines represent the mean of 100 bootstrap predictions and shaded areas represent 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Relative mean ( $\pm$ SD) percentage contribution of variables are provided in parentheses

differed slightly, ranging from +1.3% for chlorophyll *a* and  $-4.2\%$  for Shellhash cover (Table 4).

A larger change in variable importance was observed with the inclusion of the *Macomona* patch variable (10.5% importance, Env + species + patch, Table 4). The importance of the variable seagrass increased +4.0%, whereas all other variables declined in importance (Table 4). There was a positive relationship between the abundance of *Austrovenus* and *Macomona* patch at lower abundances (0.0–0.6,  $\log(\text{abundance } Macomona) + 1$ ) followed by a sharp drop at approx.

0.6 and a more gradual decline along the gradient of *Macomona* patch (Figure 3).

## 4 | DISCUSSION

The abundance of two ecologically important intertidal species – *Austrovenus* and *Macomona* – was predicted using models fitted with environmental variables only, and models fitted with biotic

TABLE 3 Variable contributions (%) explaining the variation in bootstrapped BRT models fitted with log + 1 transformed abundance estimates of *Macomona*

Variable	Env	Env + species	Difference (%)	Env + species + patch	Difference (%)
Dev (eval)	0.18 ( $\pm 0.05$ )	0.37 ( $\pm 0.04$ )		0.38 ( $\pm 0.05$ )	
<i>Austrovenus</i>		36.7 ( $\pm 3.1$ )	36.7	25.7 ( $\pm 3.4$ )	-11.0
<i>Austrovenus</i> patch				14.1 ( $\pm 2.6$ )	14.1
Chla	20.2 ( $\pm 3.4$ )	10.6 ( $\pm 3.0$ )	-9.6	9.3 ( $\pm 1.7$ )	-1.3
LOI	11.4 ( $\pm 2.1$ )	6.1 ( $\pm 1.5$ )	-5.3	5 ( $\pm 1.1$ )	-1.1
Mud	27.1 ( $\pm 4.2$ )	11.9 ( $\pm 1.5$ )	-15.2	12 ( $\pm 1.4$ )	0.1
Pheophytin	8.9 ( $\pm 1.6$ )	5.3 ( $\pm 1.2$ )	-3.6	4.2 ( $\pm 0.7$ )	-1.1
Estuary	20.3 ( $\pm 6.4$ )	20.8 ( $\pm 2.8$ )	0.5	21.3 ( $\pm 3$ )	0.5
Seagrass	8.6 ( $\pm 2.5$ )	5.5 ( $\pm 1.3$ )	-3.1	5.4 ( $\pm 0.3$ )	-0.1
Shellhash cover	3.5 ( $\pm 1.2$ )	3.2 ( $\pm 0.9$ )	-0.3	3 ( $\pm 1$ )	-0.2

Note: Mean ( $\pm$ SD) variable contributions are shown for BRT models including environmental predictors only (Env), environmental predictors and abundance of co-occurring *Austrovenus stutchburyi* (Env + species) and environmental predictors, abundance of co-occurring *Austrovenus stutchburyi* and abundance of *Austrovenus stutchburyi* from inferred patches (Env + species + patch). Difference columns represent the difference in variable contributions between the two preceding models where positive differences are shown in blue and negative differences are shown in red. Deviance explained calculated on withheld evaluation (Dev (Eval)) is provided for context.

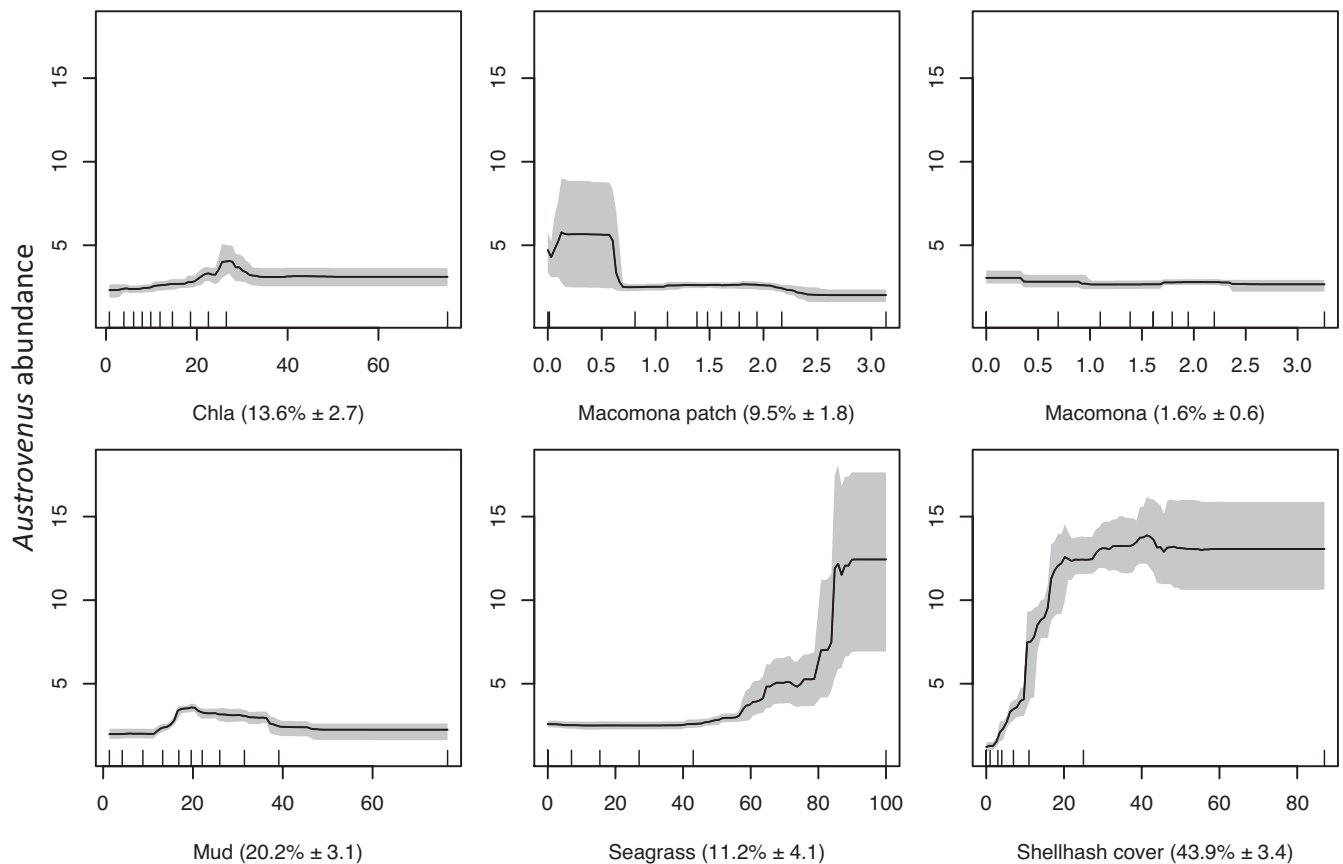


FIGURE 3 Partial dependence plots showing the relationships between predictor variables and abundance of *Austrovenus* modelled using bootstrapped BRT (environmental and biotic predictors (Env + species + patch)). Black lines represent the mean of 100 bootstrap predictions and shaded areas represent 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Relative mean ( $\pm$ SD) percentage contribution of variables are provided in parentheses

variables. We found that the inclusion of biotic variables confirmed asymmetrical interactions between species, with a strong, non-linear effect of the abundance of *Austrovenus* on *Macomona* but only

a weak effect of *Macomona* on *Austrovenus*. The inclusion of biotic variables improved the model fit metrics for both species, markedly so for *Macomona*. Whilst some recent work has investigated

TABLE 4 Variable contributions for the bootstrapped BRT models fitted with log + 1 transformed abundance estimates of *Austrovenus*

Variable	Env	Env + species	Difference (%)	Env + species + patch	Difference (%)
Dev (Eval)	0.40 ( $\pm$ 0.08)	0.41 ( $\pm$ 0.06)		0.46 ( $\pm$ 0.06)	
Chla	14.5 ( $\pm$ 2.9)	15.8 ( $\pm$ 2.9)	1.3	13.3 ( $\pm$ 2.9)	-2.5
<i>Macomona</i> patch				10.5 ( $\pm$ 2)	10.5
<i>Macomona</i>		5 ( $\pm$ 0.9)	5.0	1.3 ( $\pm$ 0.6)	-3.7
Mud	24 ( $\pm$ 3.1)	24.4 ( $\pm$ 2.2)	0.4	20.4 ( $\pm$ 2.5)	-4.0
Seagrass	11.8 ( $\pm$ 3.4)	9.4 ( $\pm$ 4)	-2.4	13.4 ( $\pm$ 1.5)	4.0
Shellhash cover	49.6 ( $\pm$ 4.3)	45.4 ( $\pm$ 3.7)	-4.2	41.1 ( $\pm$ 4.4)	-4.3

Note: Mean ( $\pm$ SD) variable contributions are shown for BRT models including environmental predictors only (Env), environmental predictors and abundance of co-occurring *Macomona* (Env + species) and environmental predictors, abundance of co-occurring *Macomona* and abundance of *Macomona* from inferred patches. Difference columns represent the difference in variable contributions between the two preceding models where positive differences are shown in blue and negative differences are shown in red. Deviance explained calculated on withheld evaluation (Dev (Eval)) is provided for context.

the species co-occurrences on intertidal soft sediments (including *Macomona* and *Austrovenus*, e.g. Kraan, Thrush, et al., 2020), here we provide novel insight into the non-linear, and asymmetrical relationships between two key intertidal species and importantly the effect that these may have on our ability to estimate abundance.

#### 4.1 | Incorporation of biotic variables in species distribution modelling

Most marine SDM studies have focused on predicting species occurrence, whereas abundance is much less frequently predicted (Melo-Merino et al., 2020). A comparison of the spatial predictions of occurrence compared to density (abundance/area) of *Austrovenus* revealed that density estimates were much more constrained than the broader and less informative predictions of occurrence (Rullens et al., 2021). This has important implications for conservation planning because the more extensive area identified from the occurrence models may not have the highest densities and therefore may not be the best areas for conservation (Rullens et al., 2021).

There are clear examples of how biotic interactions can affect species' distributions (Godsoe, Jankowski, et al., 2017); however, the importance of such interactions can vary between taxa and may be difficult to empirically determine with SDMs. That is, the use of biotic variables in SDMs (i.e. the co-occurrence of another taxa, as was the case here) may not correctly identify the direction of interactions (Ovaskainen et al., 2017), or when assigned to many taxa simultaneously, may find correlations between taxa that are not supported, or are contradictory to experimental findings (Barner et al., 2018). The latter point may occur when relationships assigned to biotic variables, are in fact residual spatial structures that could be attributed to missing co-variables (whether environmental or biotic) (Ovaskainen & Abrego, 2020; Poggiato et al., 2021). Finally, the overlapping and sometimes causal links between biotic and environmental interactions at different scales may mean that SDMs may not correctly attribute causality to biotic interactions (e.g. as shown in a simulation study by Godsoe,

Franklin, et al. (2017) who found that SDMs can miss substantial effects of biotic interactions).

Despite these shortcomings, the inclusion of biotic variables, when supported by ecological knowledge (as was the case here), can provide valuable insight into possible biotic interactions affecting species' distributions at different scales (Anderson, 2017). Given that estimating biotic interactions from experimental data may be limited in spatial extent, investigating apparent interactions based on observations in the field may be a viable method for testing ecological hypotheses (Leathwick & Austin, 2001; Soberón, 2007). Here, and in other work (Anderson, 2017), we found an increase in predictive power of SDMs for both species fitted with biotic variables compared to those fitted with environmental variables alone, markedly so for *Macomona*. That is, the overall deviance explained by the models increased, the correlation of predicted vs observed abundance data increased and the variability in these measures decreased (the models were more stable) with the inclusion of biotic variables. Although the overall shape of the abundance response curves remained similar between Env only models and those with the biotic variables, there were some localized differences. For example, in the *Macomona* abundance models, although the response curve of the mud content variable was visually similar in the Env model compared to the Env + species + patch models, in the former, mud content would have a far greater influence on the prediction of *Macomona* (most important variable: 27.1%). Mud content is a key determinant of *Macomona*'s fundamental niche, but the importance of mud content decreased in the Env + *Austrovenus* + *Macomona* patch model. This is probably because *Austrovenus* abundance and mud content are correlated so that when adding *Austrovenus* abundance into the model it down weights the effects of mud (which was likely acting as a proxy). The combination of the improvement in model performance and the changes in variable influence with the inclusion of biotic variables is of particular importance when predicting into unsampled space (i.e. predicting the distribution of species into the future, Araújo & Luoto, 2007). For example, if mud content were predicted to increase in the future (as may be the case as a result of different land-based activities see,

e.g. in Hillman et al., 2020) the Env model may provide overly pessimistic predictions of *Macomona* abundance because higher mud content, although preferred/tolerated by *Austrovenus*, does not have a large effect on *Austrovenus* abundance. Given the interplay between biotic and environmental interactions highlighted here, exploration of abundance patterns of *Austrovenus* and *Macomona* under possible future scenarios using conditional predictions (e.g. as described in Kissling et al., 2012; Ovaskainen & Abrego, 2020) would be of interest.

Whilst multiple different species are likely to interact with *Austrovenus* and *Macomona* to influence their distributions (e.g. see Kraan, Thrush, et al., 2020), there is generally limited information on the shape of these responses. We therefore focused on the interactions between these two key species that are underpinned by an extensive ecological literature (Thrush et al., 1996; e.g. Hewitt, Legendre, et al., 1997; Thrush et al., 2000). However, we acknowledge that by taking an approach where relationships were only investigated between two species, we do not account for how these may change with the inclusion of other species (Dormann et al., 2018). For example, the interactions observed between *Austrovenus* and *Macomona* may be inhibitory (see following section) but it is not clear how these interactions may be affected, by, for example, a third taxa such as a predator. To elucidate some of these questions, other techniques, such as jSDMs, may be better (Ovaskainen et al., 2017) but care must still be taken in the interpretation of correlations between study taxa and biotic variables as evidence of biotic interactions. jSDMs are a recently developed suite of models and to the authors' knowledge do not account for asymmetrical relationships. Further, it can be difficult to investigate non-linear relationships within these models if the shape of the responses is not known *a priori*. In contrast, the ability to identify and explore non-linear relationships are one of the strengths of the flexible machine learning methods used here (Leathwick et al., 2006).

Another component that is not well explored in SDMs and jSDMs is the role of scale. Indeed, biotic interactions are often only attributed to small scale processes, although biotic relationships may occur and be expressed at a variety of scales (Thrush et al., 2021). We included a scale factor (density within the surrounding 70 m cf density in 13 cm) and observed strong differences between species in the relative importance of these measures. When predicting *Macomona* abundance, the density of *Austrovenus* at the small scale was roughly twice as important as the larger patch. The inclusion of the larger patch factor barely changed the explanatory power of the model, splitting the explanatory power of the small-scale *Austrovenus* density, and also removing small amounts of the explanatory power of chlorophyll and seagrass. However, when predicting *Austrovenus* abundance, density of *Macomona* at the small scale barely increased the explanatory power of the model, compared to a 5% increase when the larger patch density was included as a predictor. The large patch density explained 10% of the total explained, removing the effect of the small-scale density, reducing the effect of shell-hash, mud and chlorophyll *a*, and increasing the effect of seagrass.

## 4.2 | Ecological considerations

Our models showed strong non-linear effects of *Austrovenus* on *Macomona*, where low to moderate densities of *Austrovenus* resulted in sharp declines in *Macomona* abundance. The low threshold of this effect means that even the inclusion of *Austrovenus* presence may already be a good indicator of *Macomona* abundance; this could potentially have implications for monitoring where predictive models of *Macomona* could vastly be improved with the inclusion of rapidly collected (hand raked) field data.

The biophysical feedbacks between *Austrovenus* and their physical environment have been extensively explored and provide a basis to explain the non-linear effects that *Austrovenus* has on *Macomona*. Both *Austrovenus* and *Macomona* stimulate nutrient regeneration in the sediments and enhance the growth and productivity of microphytobenthos (Jones et al., 2011; Sandwell et al., 2009; Thrush et al., 2006; Woodin et al., 2016), the primary food source of *Macomona*. However, at high densities of *Austrovenus* (i.e. when they occur with little to no distance between shells in the surface sediments), this may preclude the siphons of *Macomona* from reaching the sediment surface. Thus, there may be a density-dependent switch between a positive and negative inhibitory interaction between the two species. Similarly to studies in terrestrial, freshwater and marine systems, here we show that the direction and magnitude of species interactions varied along spatial and abiotic gradients (Chamberlain et al., 2014). In our models, *Macomona* abundance was not a good predictor of *Austrovenus* distribution and this confirms an asymmetric interaction between the two species. This asymmetry may in part be caused by the relative differences in the "size" of the fundamental niches for each species. *Austrovenus* has a broader fundamental niche than *Macomona* and therefore is more likely to be influenced by the physical environment than by the abundance of *Macomona* which only occurs in a subset of the fundamental niche of *Austrovenus*. Further, *Austrovenus* create structure in the sediment and modify their physical environment, which may create positive feedbacks with the environment that more strongly drive their distributions than their interactions with other species. For example, shellhash was a good predictor of *Austrovenus* abundance, likely because the *Austrovenus* themselves are generating the shellhash. The cyclical positive feedback of shellhash creation is not well captured in (temporally static) correlative models. Future investigation may benefit from accounting for these cyclical processes, for example by using Structural Equation Models which can account for positive feedback (e.g. as in Thrush et al., 2012).

## 5 | CONCLUSION

Here, we have applied ecological knowledge of two key species generated from experimental research and integrated this knowledge into quantitative SDMs. We show that an understanding of the ecology of species and a consideration of scale is critical for increasing our capacity to predict meaningful species distributions although

care is required when interpreting correlations as relationships. This selective choice of biotic and environmental variables is likely to produce parsimonious models which are more likely to be accurate for prediction into unsampled space.

## ACKNOWLEDGEMENTS

The authors thank Conrad Pilditch, Joanne Ellis and Casper Kraan for advice and discussions on the manuscript and Grady Petersen for helping design Figure 1. Data collection and processing was supported by the Marsden Fund of the Royal Society of New Zealand (NIW-1102), and a Marie-Curie fellowship (Grant700796). The analysis and writing of this manuscript were funded by the National Institute of Water and Atmospheric Research Coast and Oceans SSIF funding (projects COME2001 and COME2101).

## CONFLICT OF INTEREST

The authors declare no conflicts of interest.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/ddi.13546>.

## DATA AVAILABILITY STATEMENT

All data used in this research are described and available from Kraan, Thrush, et al. (2020). All R code used in the analysis is available in the supplementary materials.

## ORCID

Fabrice Stephenson  <https://orcid.org/0000-0002-9500-5204>

## REFERENCES

- Anderson, R. P. (2017). When and how should biotic interactions be considered in models of species niches and distributions? *Journal of Biogeography*, 44, 8–17. <https://doi.org/10.1111/jbi.12825>
- Araújo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, 33, 1677–1688. <https://doi.org/10.1111/j.1365-2699.2006.01584.x>
- Araújo, M. B., & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, 16, 743–753. <https://doi.org/10.1111/j.1466-8238.2007.00359.x>
- Barner, A. K., Coblenz, K. E., Hacker, S. D., & Menge, B. A. (2018). Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. *Ecology*, 99, 557–566. <https://doi.org/10.1002/ecy.2133>
- Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, 23, 1050–1063. <https://doi.org/10.1111/ele.13525>
- Chamberlain, S. A., Bronstein, J. L., & Rudgers, J. A. (2014). How context dependent are species interactions? *Ecology Letters*, 17, 881–890. <https://doi.org/10.1111/ele.12279>
- Compton, T. J., Morrison, M. A., Leathwick, J. R., & Carabines, G. D. (2012). Ontogenetic habitat associations of a demersal fish species, *Pagrus auratus*, identified using boosted regression trees. *Marine Ecology Progress Series*, 462, 219–230. <https://doi.org/10.3354/meps09790>
- Dormann, C. F., Bobrowski, M., Dehling, D. M., Harris, D. J., Hartig, F., Lischke, H., Moretti, M. D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S. I., Sheppard, C. S., Steinbauer, M. J., Zeuss, D., & Kraan, C. (2018). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography*, 27, 1004–1016. <https://doi.org/10.1111/geb.12759>
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münkemüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Douglas, E. J., Lohrer, A. M., & Pilditch, C. A. (2019). Biodiversity breakpoints along stress gradients in estuaries and associated shifts in ecosystem interactions. *Scientific Reports*, 9, 1–11. <https://doi.org/10.1038/s41598-019-54192-0>
- Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., ... E. Zimmermann, N. (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., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1, 330–342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Ellis, J. I., Clark, D., Atalah, J., Jiang, W., Taiapa, C., Patterson, M., Sinner, J., & Hewitt, J. (2017). Multiple stressor effects on marine infauna: responses of estuarine taxa and functional traits to sedimentation, nutrient and metal loading. *Scientific Reports*, 7, 1–16. <https://doi.org/10.1038/s41598-017-12323-5>
- 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>
- Friedman, J. H., & Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, 22, 1365–1381. <https://doi.org/10.1002/sim.1501>
- Godsoe, W., Franklin, J., & Blanchet, F. G. (2017). Effects of biotic interactions on modeled species' distribution can be masked by environmental gradients. *Ecology and Evolution*, 7, 654–664. <https://doi.org/10.1002/ece3.2657>
- Godsoe, W., Jankowski, J., Holt, R. D., & Gravel, D. (2017). Integrating biogeography with contemporary Niche Theory. *Trends in Ecology & Evolution*, 32, 488–499. <https://doi.org/10.1016/j.tree.2017.03.008>
- Harris, D. J. (2016). Inferring species interactions from co-occurrence data with Markov networks. *Ecology*, 97, 3308–3314. <https://doi.org/10.1002/ecy.1605>
- He, Q., & Bertness, M. D. (2014). Extreme stresses, niches, and positive species interactions along stress gradients. *Ecology*, 95, 1437–1443. <https://doi.org/10.1890/13-2226.1>
- Hewitt, J. E., Anderson, M. J., Hickey, C. W., Kelly, S., & Thrush, S. F. (2009). Enhancing the ecological significance of sediment contamination guidelines through integration with community analysis. *Environmental Science & Technology*, 43, 2118–2123. <https://doi.org/10.1021/es802175k>
- Hewitt, J., Legendre, P., McArdle, B., Thrush, S., Bellehumeur, C., & Lawrie, S. (1997). Identifying relationships between adult and juvenile bivalves at different spatial scales. *Journal of Experimental Marine Biology and Ecology*, 216, 77–98. [https://doi.org/10.1016/S0022-0981\(97\)00091-9](https://doi.org/10.1016/S0022-0981(97)00091-9)

- Hewitt, J., Thrush, S., Cummings, V., & Pridmore, R. (1997). Matching patterns with processes: predicting the effect of size and mobility on the spatial distributions of the bivalves *Macomona liliiana* and *Austrovenus stutchburyi*. *Oceanographic Literature Review*, 1, 48. <https://doi.org/10.3354/meps135057>
- Hewitt, J. E., Thrush, S. F., Halliday, J., & Duffy, C. (2005). The importance of small-scale habitat structure for maintaining beta diversity. *Ecology*, 86, 1619–1626. <https://doi.org/10.1890/04-1099>
- Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). *dismo: Species Distribution Modeling R package version 1.1-4*. <https://CRAN.R-project.org/package=dismo>
- Hillman, J. R., Stephenson, F., Thrush, S. F., & Lundquist, C. J. (2020). Investigating changes in estuarine ecosystem functioning under future scenarios. *Ecological Applications*, 30, 2090. <https://doi.org/10.1002/eap.2090>
- Hutchinson, G. E. (1978). *An introduction to population ecology*. Yale University Press.
- Jones, H. F., Pilditch, C. A., Bruesewitz, D. A., & Lohrer, A. M. (2011). Sedimentary environment influences the effect of an infaunal suspension feeding bivalve on estuarine ecosystem function. *PLoS One*, 6, e27065. <https://doi.org/10.1371/journal.pone.0027065>
- Kissling, W. D., Dormann, C. F., Groeneveld, J., Hickler, T., Kühn, I., McInerney, G. J., Montoya, J. M., Römermann, C., Schiffers, K., Schurr, F. M., Singer, A., Svenning, J.-C., Zimmermann, N. E., & O'Hara, R. B. (2012). Towards novel approaches to modelling biotic interactions in multispecies assemblages at large spatial extents. *Journal of Biogeography*, 39, 2163–2178. <https://doi.org/10.1111/j.1365-2699.2011.02663.x>
- Kraan, C., Dormann, C. F., Greenfield, B. L., & Thrush, S. F. (2015). Cross-scale variation in biodiversity-environment links illustrated by coastal sandflat communities. *PLoS One*, 10, e0142411. <https://doi.org/10.1371/journal.pone.0142411>
- Kraan, C., Greenfield, B. L., & Thrush, S. F. (2020). Multi-scale data on intertidal macrobenthic biodiversity and environmental features in three New Zealand harbours. *Earth System Science Data*, 12, 293–297. <https://doi.org/10.5194/essd-12-293-2020>
- Kraan, C., Thrush, S. F., & Dormann, C. F. (2020). Co-occurrence patterns and the large-scale spatial structure of benthic communities in seagrass meadows and bare sand. *BMC Ecology*, 20, 37. <https://doi.org/10.1186/s12898-020-00308-4>
- Leathwick, J. R., & Austin, M. P. (2001). Competitive interactions between tree species in New Zealand's old-growth indigenous forests. *Ecology*, 82, 2560–2573.
- 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>
- Lohrer, A. M., Townsend, M., Hales, S. F., Rodil, I. F., Cartner, K., Pratt, D. R., & Hewitt, J. E. (2016). Influence of New Zealand cockles (*Austrovenus stutchburyi*) on primary productivity in sandflat-seagrass (*Zostera muelleri*) ecotones. *Estuarine, Coastal and Shelf Science*, 181, 238–248. <https://doi.org/10.1016/j.ecss.2016.08.045>
- Lohrer, A. M., Townsend, M., Rodil, I. F., Hewitt, J. E., & Thrush, S. F. (2012). Detecting shifts in ecosystem functioning: The decoupling of fundamental relationships with increased pollutant stress on sandflats. *Marine Pollution Bulletin*, 64, 2761–2769. <https://doi.org/10.1016/j.marpolbul.2012.09.012>
- Louthan, A. M., Doak, D. F., & Angert, A. L. (2015). Where and when do species interactions set range limits? *Trends in Ecology & Evolution*, 30, 780–792. <https://doi.org/10.1016/j.tree.2015.09.011>
- Maire, O., Merchant, J. N., Bulling, M., Teal, L. R., Grémare, A., Duchêne, J. C., & Solan, M. (2010). Indirect effects of non-lethal predation on bivalve activity and sediment reworking. *Journal of Experimental Marine Biology and Ecology*, 395, 30–36. <https://doi.org/10.1016/j.jembe.2010.08.004>
- Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, 405, 243. <https://doi.org/10.1038/35012251>
- Mason, T. H., Stephens, P. A., Apollonio, M., & Willis, S. G. (2014). Predicting potential responses to future climate in an alpine ungulate: interspecific interactions exceed climate effects. *Global Change Biology*, 20, 3872–3882. <https://doi.org/10.1111/gcb.12641>
- Melo-Merino, S. M., Reyes-Bonilla, H., & Lira-Noriega, A. (2020). Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecological Modelling*, 415, 108837. <https://doi.org/10.1016/j.ecolmodel.2019.108837>
- Ovaskainen, O., & Abrego, N. (2020). *Joint species distribution modelling: With applications in R*. Cambridge University Press.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20, 561–576. <https://doi.org/10.1111/ele.12757>
- Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J. S., & Thuiller, W. (2021). On the interpretations of joint modeling in Community Ecology. *Trends in Ecology & Evolution*, 36(5), 391–401. <https://doi.org/10.1016/j.tree.2021.01.002>
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Ridgeway, G. (2007). *Generalized Boosted Models: A guide to the gbm package*.
- Rullens, V., Lohrer, A. M., Townsend, M., & Pilditch, C. A. (2019). Ecological mechanisms underpinning ecosystem service bundles in marine environments – A case study for shellfish. *Frontiers in Marine Science*, 6, 409.
- Rullens, V., Stephenson, F., Lohrer, A. M., Townsend, M., & Pilditch, C. A. (2021). Combined species occurrence and density predictions to improve marine spatial management. *Ocean & Coastal Management*, 209, 105697. <https://doi.org/10.1016/j.ocecoaman.2021.105697>
- Sandwell, D. R., Pilditch, C. A., & Lohrer, A. M. (2009). Density dependent effects of an infaunal suspension-feeding bivalve (*Austrovenus stutchburyi*) on sandflat nutrient fluxes and microphytobenthic productivity. *Journal of Experimental Marine Biology and Ecology*, 373, 16–25. <https://doi.org/10.1016/j.jembe.2009.02.015>
- Soberón, J. (2007). Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, 10, 1115–1123. <https://doi.org/10.1111/j.1461-0248.2007.01107.x>
- Stachowicz, J. (2012). Niche expansion by positive interactions: Realizing the fundamentals. A comment on Rodriguez-Cabal et al. *Ideas in Ecology and Evolution*, 5, 42–43.
- Thrush, S. F., Hewitt, J. E., Cummings, V. J., Green, M. O., Funnell, G. A., & Wilkinson, M. R. (2000). The generality of field experiments: Interactions between local and broad-scale processes. *Ecology*, 81, 399–415.
- Thrush, S. F., Hewitt, J. E., Gibbs, M., Lundquist, C., & Norkko, A. (2006). Functional role of large organisms in intertidal communities: Community effects and ecosystem function. *Ecosystems*, 9, 1029–1040. <https://doi.org/10.1007/s10021-005-0068-8>
- Thrush, S., Hewitt, J., & Lohrer, A. (2012). Interaction networks in coastal soft-sediments highlight the potential for change in ecological resilience. *Ecological Applications*, 22, 1213–1223. <https://doi.org/10.1890/11-1403.1>

- Thrush, S. F., Hewitt, J. E., Norkko, A., Nicholls, P. E., Funnell, G. A., & Ellis, J. I. (2003). Habitat change in estuaries: predicting broad-scale responses of intertidal macrofauna to sediment mud content. *Marine Ecology Progress Series*, 263, 101–112. <https://doi.org/10.3354/meps263101>
- Thrush, S. F., Hewitt, J. E., Pilditch, C. A., & Norkko, A. (2021). Design and the philosophy of sampling. In *Ecology of coastal marine sediments*. Oxford University Press. <https://doi.org/10.1093/oso/9780198804765.001.0001>
- Thrush, S. F., Hewitt, J. E., Pridmore, R. D., & Cummings, V. J. (1996). Adult/juvenile interactions of infaunal bivalves: Contrasting outcomes in different habitats. *Marine Ecology Progress Series*, 132, 83–92. <https://doi.org/10.3354/meps132083>
- Weber, M. M., Stevens, R. D., Diniz-Filho, J. A. F., & Grelle, C. E. V. (2017). Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. *Ecography*, 40, 817–828. <https://doi.org/10.1111/ecog.02125>
- Woodin, S. A., Volkenborn, N., Pilditch, C. A., Lohrer, A. M., Wetthey, D. S., Hewitt, J. E., & Thrush, S. F. (2016). Same pattern, different mechanism: Locking onto the role of key species in seafloor ecosystem process. *Scientific Reports*, 6, 26678.
- Zurell, D., Franklin, J., König, C., Bouchet, P. J., Dormann, C. F., Elith, J., Fandos, G., Feng, X., Guillera-Aroita, G., Guisan, A., Lahoz-Monfort, J. J., Leitão, P. J., Park, D. S., Peterson, A. T., Rapacciuolo, G., Schmatz, D. R., Schröder, B., Serra-Diaz, J. M., Thuiller, W., ... Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography*, 43, 1261–1277. <https://doi.org/10.1111/ecog.04960>

## BIOSKETCH

**Fabrice Stephenson** is a quantitative marine ecologist with interests in modelling spatial patterns of biodiversity for use in conservation and marine spatial planning. Rebecca V. Gladstone-Gallagher, Richard H. Bulmer, Simon F. Thrush and Judi E. Hewitt are benthic ecologists with interests in explaining biodiversity patterns at a variety of scales.

Author contributions: F.S. and J.E.H. conceived the ideas with input from all authors. F.S. and J.E.H. led the analysis. F.S. and R.V.G.G. led the writing of the manuscript with input from all authors.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Stephenson, F., Gladstone-Gallagher, R. V., Bulmer, R. H., Thrush, S. F., & Hewitt, J. E. (2022). Inclusion of biotic variables improves predictions of environmental niche models. *Diversity and Distributions*, 28, 1373–1390. <https://doi.org/10.1111/ddi.13546>

## APPENDIX 1

## ODMAP (Overview, Data, Model, Assessment and Prediction) standardized protocol

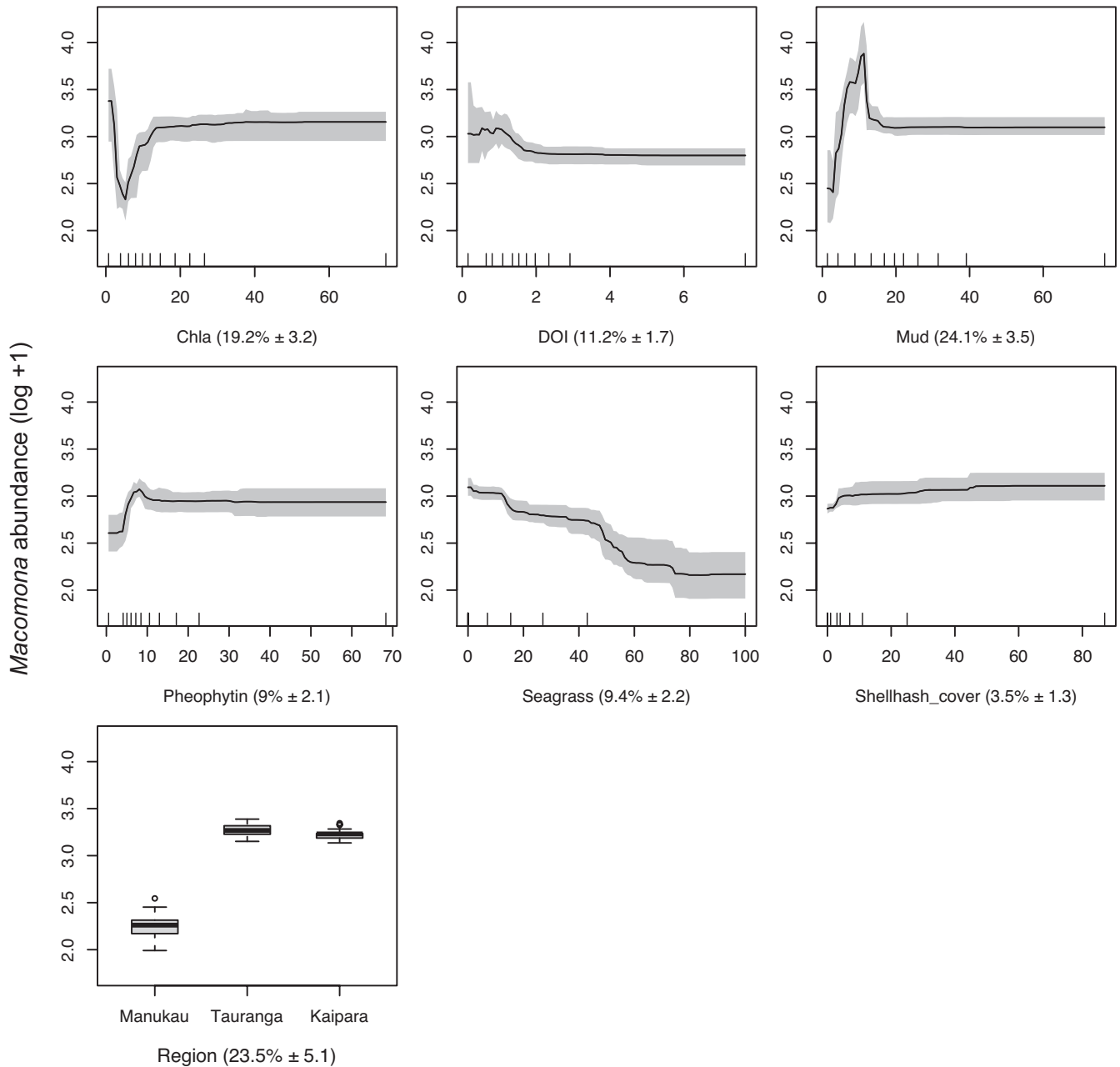
TABLE A1 ODMAP for journal article: Inclusion of biotic variables improve predictions of environmental niche models

ODMAP element	Contents
Overview	
<i>Authorship</i>	<ul style="list-style-type: none"> <li>• Authors: Fabrice Stephenson, Rebecca V. Gladstone-Gallagher, Richard H Bulmer, Simon F Thrush and Judi E Hewitt</li> <li>• Contact email: <a href="mailto:fabrice.stephenson@waikato.ac.nz">fabrice.stephenson@waikato.ac.nz</a></li> <li>• Title: Inclusion of biotic variables improve predictions of environmental niche models.</li> <li>• DOI:</li> </ul>
<i>Model objective</i>	<ul style="list-style-type: none"> <li>• SDM objective: ecological inference/explanation. To explore the possible influence of biotic interactions in explaining patterns of species' distribution, abundance and explanatory power of SDMs, we modelled the distributions of two intertidal shellfish (<i>Austrovenus stutchburyi</i> and <i>Macomona liliana</i>) with and without biotic variables.</li> <li>• Target outputs: Visualisations of relationships between species' abundance and environmental and biotic variables (provided as partial dependence)</li> </ul>
<i>Taxon</i>	<i>Austrovenus stutchburyi</i> and <i>Macomona liliana</i>
<i>Location</i>	New Zealand
<i>Scale of analysis</i>	<ul style="list-style-type: none"> <li>• Spatial Extent (Lon / Lat): 36.4°–37.5°S, 174.3°–176.0°W</li> <li>• Spatial resolution: 1 m</li> <li>• Temporal extent/time period: Austral summer 2012</li> </ul>
<i>Biodiversity data overview</i>	<ul style="list-style-type: none"> <li>• Observation type: field survey</li> <li>• Response/data type: counts/macrofauna core (13 cm diameter, 20 cm deep)</li> </ul>
<i>Types of predictors</i>	Environmental and biotic variables
<i>Conceptual model/ Hypotheses</i>	We used key environmental variables that are known to shape our study species' distribution in intertidal sand flats (Hewitt et al., 2005; Thrush et al., 2003). This included chlorophyll <i>a</i> and pheophytin measurements, sediment grain-size distributions, organic content of the sediment and visible sandflat parameters (e.g. the coverage of seagrass, shellhash and bare sand) estimated within each photograph of the sample site prior to macrofauna sampling. In addition to environmental variables, we explored the influence of biotic variables in explaining patterns of each of our study taxa. This included the co-occurring abundance of the study taxa not being modelled (i.e. <i>Macomona</i> or <i>Austrovenus</i> ), as well as a spatially smoothed representation of this data used to represent a "patch," that is. an area of species aggregation defined using Inverse Distance Weighting (IDW) of the co-occurring abundance sample locations with a circular search radius of 70 m diameter (smoothing parameter based on prior estimates of patch size in New Zealand soft sediment habitats – Hewitt, Legendre, et al. (1997)
<i>Assumptions</i>	<ul style="list-style-type: none"> <li>• #1: Relevant ecological drivers (or proxies) of species distributions are included.</li> <li>• #2: Detectability does not change across habitat gradients.</li> <li>• #3: Species are at equilibrium with their environment.</li> <li>• #4: Sampling is adequate and representative (and any biases are accounted for/corrected).</li> </ul>
<i>SDM algorithms</i>	<ul style="list-style-type: none"> <li>• Algorithm: Boosted Regression Trees. Chosen due to competitive performance (machine learning algorithm) and ability to easily incorporate non-linear relationships.</li> <li>• Model complexity: we allowed 3-way interactions (interaction depth =3) to account for important interactions between our study taxa and environmental and biotic variables</li> <li>• Model averaging: model averaging (<math>n = 100</math> bootstrap replicates)</li> </ul>
<i>Model workflow</i>	<p>The relationship between environmental and biotic variables on the abundance of our two species was investigated using Boosted Regression Trees (BRTs) with increasing model complexity. Firstly, BRT models using environmental variables were fitted to each species (referred throughout as Env models). Secondly, in order to explore the effect of including biotic variables, BRT models using environmental variables and the co-occurring abundance of the study taxa not being modelled were fitted (referred throughout as Env + species models). Finally, BRT models using environmental variables, the co-occurring abundance and the estimated abundance of the "species' patch" of the study taxa not being modelled were fitted (referred throughout as Env +species + patch models).</p> <p>Environmental predictor variable selection was undertaken using a simplification process whereby an initial model was fitted with all the environmental variables which were then removed from the model, one at a time, using the "simplify" function (Elith et al., 2006). This simplification process firstly assesses the relative contributions of each variable in terms of deviance explained, with the lowest contributing variables removed from the model, before the model is refitted with the remaining environmental variables (this is somewhat analogous to a stepwise backward selection in linear modelling). The final models were created by refitting the model with a reduced variable set that balanced the deviance explained with a reasonable number of predictor variables</p> <p>All levels of co-linearity between predictor variables were considered acceptable (Pearson correlation &lt;0.8) for tree-based machine learning methods (Dormann et al., 2013; Elith et al., 2010).</p> <p>BRT models were assessed using cross-validated measures of model performance, including the deviance explained and the Pearson correlation between predicted and observed abundance.</p>

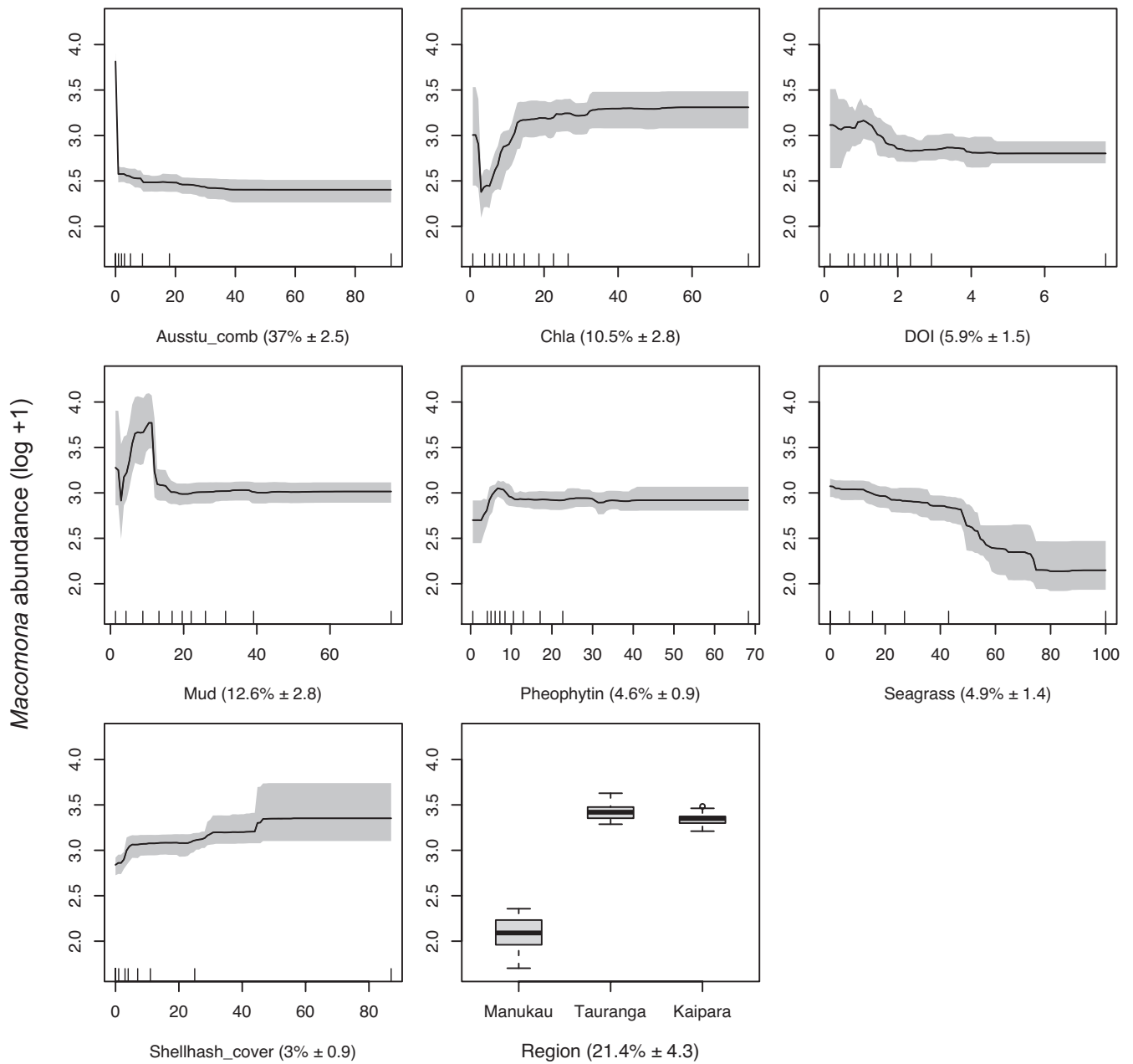
TABLE A1 (Continued)

ODMAP element	Contents
Software	<ul style="list-style-type: none"> <li>• Software: R (no version specified). Dismo package (v1.1–4)</li> <li>• Availability of code: <a href="#">Appendix 1</a>, Section 1.3 of <i>Inclusion of biotic variables improve predictions of environmental niche models</i>.</li> <li>• Data availability: <a href="https://doi.org/10.5194/essd-12-293-2020">https://doi.org/10.5194/essd-12-293-2020</a> (Kraan, Thrush, et al., 2020)</li> </ul>
Data	
Biodiversity data	<ul style="list-style-type: none"> <li>• Taxon names: <i>Austrovenus stutchburyi</i> and <i>Macomona liliana</i></li> <li>• Taxonomic reference system: We follow the taxonomy of World Register of Marine Species (WoRMS) (<a href="https://www.marinespecies.org/">https://www.marinespecies.org/</a>)</li> <li>• Ecological level: species</li> <li>• Data sources: <a href="https://doi.org/10.5194/essd-12-293-2020">https://doi.org/10.5194/essd-12-293-2020</a></li> <li>• Sampling design: uniform spatial design was undertaken in three Harbours in New Zealand. In each harbour, macrobenthic fauna were collected and abundance recorded from 400 cores (13 cm diameter, 20 cm deep) sampled along four 1000 m transects (spaced 100 m apart) during low tide from the high- to low-water mark. Sampling points along the transects were spaced at distances of 1, 5, 10, 30, 50, 100, 500 and 1000 m (Kraan et al., 2015).</li> <li>• Sample size per taxon: <math>n = 853</math></li> <li>• Data filtering: Only samples containing at least one of either of our study taxa were retained for the analysis</li> <li>• Background data: NA – this study was focussed on abundance estimates conditional on presence</li> </ul>
Data partitioning	BRT models were bootstrapped 100 times for each study taxa. That is, a random “training” sample with a sample size equal to the number of abundance records was drawn with replacement and a model was constructed with the same settings as the original. This process was repeated 100 times, and at each iteration, predictions were made to the “evaluation” data (the remaining data not randomly selected from the abundance database) allowing model fits to be examined both on the training and evaluation data.
Predictor variables	<ul style="list-style-type: none"> <li>• Predictor variables: At each sampling point predictor variables were collected: <ul style="list-style-type: none"> <li>• <i>Environmental variables</i>: chlorophyll <i>a</i> and pheophytin measurements, sediment grain-size distributions, organic content of the sediment and visible sandflat parameters (e.g. the coverage of seagrass, shellhash and bare sand estimated within each photograph of the sample site)</li> <li>• <i>Biotic variables</i>: the co-occurring abundance of the study taxa not being modelled (i.e. <i>Macomona</i> or <i>Austrovenus</i>), as well as a spatially smoothed representation of this data used to represent a “patch,” that is an area of species aggregation defined using Inverse Distance Weighting (IDW) of the co-occurring abundance sample locations with a circular search radius of 70 m diameter</li> <li>• <i>Non-specific spatial variable</i>: “Estuary” was included as a factor to explain spatial differences in the abundance of study taxa not explained by environmental and biotic variables</li> </ul> </li> <li>• Data source: Sampled at each study site – available: <a href="https://doi.org/10.5194/essd-12-293-2020">https://doi.org/10.5194/essd-12-293-2020</a></li> </ul>
Model	
Multicollinearity	All levels of co-linearity between predictor variables were considered acceptable (Pearson correlation <0.8) for tree-based machine learning methods (Dormann et al., 2013; Elith et al., 2010)
Model settings	BRT models were parametrised following recommendations from Elith et al. (2008): with tree complexity of 3 (allowing 3-way interactions of predictor variables), a learning rate of 0.001 (parameters selected so as to fit between 2000 and 3000 trees for each species’ model), a bag fraction of 0.6 (the proportion of data drawn at random, without replacement for each tree) and random 10-fold cross evaluation.
Model estimates	The relative influence of each environmental variable in the models was the number of times it was selected for splitting, weighted by the squared improvement to the model as a result of each split (using in-bag data) (Friedman & Meulman, 2003).
Non-independence correction/ analyses	NA
Assessment	
Performance statistics	Model performance measures included the mean ( $\pm$ standard deviation of the mean) deviance explained and the Pearson correlation between predicted and observed abundance for both training data and withheld validation data (from the data partitioning) across bootstrap models ( $n = 100$ )
Plausibility check	The fitted species responses were checked for ecological plausibility through the inspection of partial dependency plots

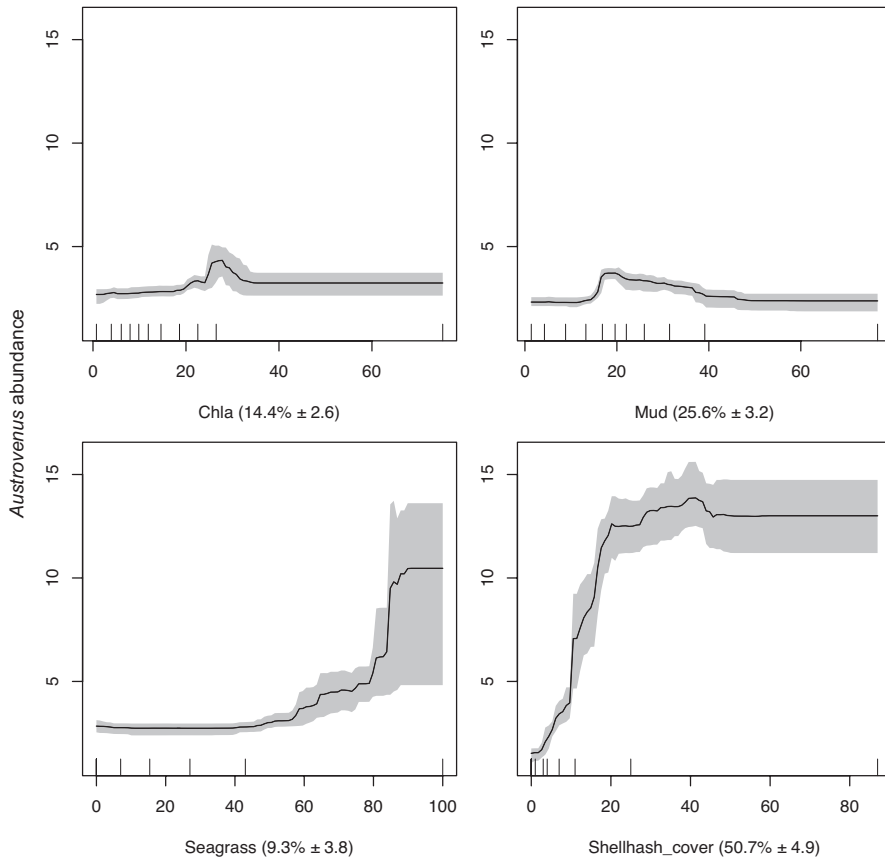
## Environmental relationships



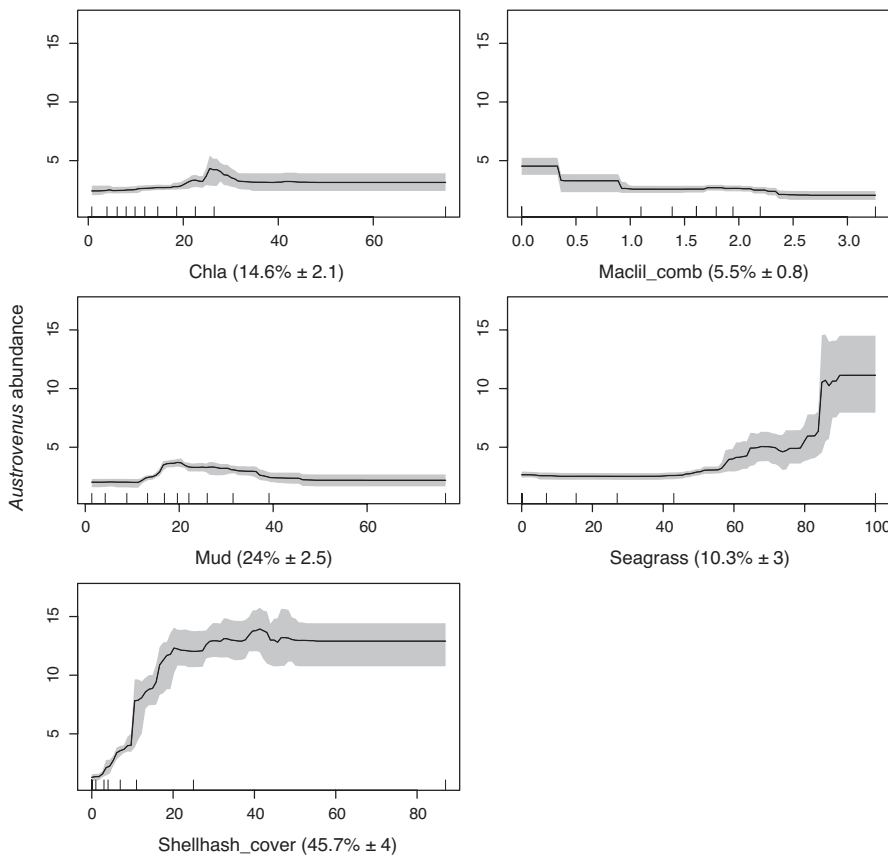
**FIGURE A1** Partial dependence plots showing the relationships between predictor variables and log +1 transformed abundance of *Macomona* modelled using bootstrapped BRT (environmental predictors only (Env)). Black lines represent the mean of 100 bootstrap predictions and shaded areas represent 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Relative mean ( $\pm$ SD) percentage contribution of variables are provided in parentheses



**FIGURE A2** Partial dependence plots showing the relationships between predictor variables and log +1 transformed abundance of *Macomona* modelled using bootstrapped BRT (environmental and biotic predictors (Env +species)). Black lines represent the mean of 100 bootstrap predictions and shaded areas represent 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Relative mean ( $\pm$ SD) percentage contribution of variables are provided in parentheses



**FIGURE A3** Partial dependence plots showing the relationships between predictor variables and abundance of *Austrovenus* modelled using bootstrapped BRT (environmental predictors only (Env)). Black lines represent the mean of 100 bootstrap predictions and shaded areas represent 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Relative mean ( $\pm$ SD) percentage contribution of variables are provided in parentheses



**FIGURE A4** Partial dependence plots showing the relationships between predictor variables and abundance of *Austrovenus* modelled using bootstrapped BRT (environmental and biotic predictors (Env +species)). Black lines represent the mean of 100 bootstrap predictions and shaded areas represent 95% prediction interval. Deciles of each environmental predictor are shown as ticks on the x-axes. Relative mean ( $\pm$ SD) percentage contribution of variables are provided in parentheses