

Essential ingredients in Joint Species Distribution Models: influence on interpretability, explanatory and predictive power

Clément Violet ¹✉, Aurélien Boyé ¹, Mathieu Chevalier ¹, Olivier Gauthier ²,
Jacques Grall ³, Martin P. Marzloff ¹

¹IFREMER, Centre de Bretagne, DYNECO LEBCO, Plouzané, France; ²Laboratoire des
Sciences de l'Environnement Marin (LEMAR) UMR 6539 CNRS UBO IRD IFREMER, Institut
Universitaire Européen de la Mer, Université de Bretagne Occidentale, Plouzané, France;
³Observatoire des Sciences de l'Univers, UMS 3113, Institut Universitaire Européen de la
Mer, Plouzané, France

✉ For correspondence:
clement.violet@ifremer.fr

Present address: IFREMER,
Centre de Bretagne, DYNECO
LEBCO, Plouzané 29280,
France.

Keywords: Community
assembly, Explanatory power,
Interpretability, Joint Species
Distribution Model, jSDM,
Model Performances,
Predictive power, Species
Distribution Model

Competing interests: The
authors declare no competing
interests.

Funding: This work benefits
from funding from IFREMER.

Abstract

1. Joint Species Distribution Models (jSDM) are increasingly used to explain and predict biodiversity patterns. jSDMs account for species co-occurrence patterns and can include phylogeny or functional traits to better capture the processes shaping communities. Yet, several factors may limit or affect the interpretability and predictive ability of jSDMs : missing abiotic predictors, omitting ecologically-important species, or increasing the number of model parameters by adding phylogeny and/or trait information.
2. We assessed how interpretability, explanatory and predictive power of jSDM varied across four alternative models focusing on 99 coastal benthic marine polychaete species: (1) a baseline jSDM with no additional information sources other than abiotic predictors and

residual co-occurrence patterns, (2) a jSDM including phylogeny alone or (3) in combination with traits data and (4) a jSDM including monitoring information related to additional species sampled alongside the target assemblage (i.e. non-target species that are not of direct interest but potentially interact with the target assemblage). The four models fitted on both presence/absence and abundance data from a regional monitoring programme were assessed using complementary metrics. We compared performance at both species- and community-level, considering multiple facets of species responses and assemblage diversity.

3. For both presence/absence and abundance data, all models displayed good and similar explanatory power but varied in their interpretability and predictive power. Considering trait data provides insights on species response along environmental gradients, which is a decisive element for model interpretability. Relative to the baseline model, predictive power increased by 26% when including data on additional species, whereas only marginal changes were detected for the two other models. These patterns are explained by changes in the species-environment relationships and residual co-occurrence patterns inferred by these models.

4. Overall, this study highlights that adequate strategy to fit jSDM depends on data at hand, modelling objective and research question. To understand observed community space-time variability, adding phylogenetic or trait information is most effective. Inclusion of non-target species is however a better strategy to predict how the target species assemblage responds to environmental changes. Importantly, we provide a comprehensive toolbox for the comparative assessment of jSDM performance.

Introduction

Community ecology aims at explaining and predicting spatio-temporal variability in species diversity (Whittaker *et al.* 2001) and coexistence (Chesson 2000). Understanding the processes that determine species distribution around the planet is a prerequisite to characterise and predict community structure and associated ecological dynamics, which is critical to mitigate the effects of global change on biodiversity and prevent the sixth mass extinction (IPBES 2019). Currently, the major challenges faced by ecologists include describing, explaining, and predicting changes in communities

52 (Tredennick *et al.* 2021) in order to inform effective management or restoration measures in a
53 rapidly changing world (Houlahan *et al.* 2017 ; Dietze *et al.* 2018 ; Brudvig & Catano 2022). Joint
54 Species Distribution Models (jSDM) are particularly well-suited tools to address these challenges,
55 whether to characterise the processes that shape observed communities (Warton *et al.* 2015 ;
56 Ovaskainen *et al.* 2017b), or to predict how communities will evolve in the future (Norberg *et al.*
57 2019 ; Pollock *et al.* 2020).

58 jSDMs are multivariate (i.e. multi-species) extensions of Species Distribution Models (SDMs),
59 which have been broadly applied over the past decades - across all terrestrial and marine realms -
60 to understand and predict both species occurrences (Elith *et al.* 2006 ; Norberg *et al.* 2019) and
61 species abundances (Howard *et al.* 2014 ; Waldock *et al.* 2022) using a set of covariates (e.g. climatic
62 variables). One advantage of jSDM relies on their explanatory power owing to their tight link with
63 the assembly rule framework (Ovaskainen *et al.* 2017b). In particular, relative to single-species
64 SDMs that only consider the abiotic niche of species (i.e. the Grinnellian niche), jSDM can theoretically
65 also account for interspecific interactions (i.e. the Eltonian niche).

66 Indeed, in jSDMs, the variability in community composition not explained by covariates is
67 captured by a residual covariance matrix representing species co-occurrence patterns potentially
68 representing biotic interactions (Ovaskainen *et al.* 2017b). This feature is highly attractive to
69 ecologists because it provides a way to disentangle the relative influence of abiotic and biotic
70 processes on biodiversity patterns (Godsoe *et al.* 2017) while also improving model's predictive
71 power (Giannini *et al.* 2013 ; Staniczenko *et al.* 2017). However, in practice, inferring and interpreting
72 residual co-occurrence patterns using jSDMs remains challenging for several reasons (Blanchet *et al.*
73 2020 ; Holt 2020).

74 First, while jSDMs have been applied to a large number of species presence/absence datasets
75 (Norberg *et al.* 2019 ; Wilkinson *et al.* 2019 ; Wilkinson *et al.* 2020), simulation studies showed that
76 co-occurrence networks inferred from such data does not necessarily provide evidence for species
77 interactions (Sander *et al.* 2017 ; Dormann *et al.* 2018 ; Blanchet *et al.* 2020) and only inform about
78 spatial and temporal associations between species (Keil *et al.* 2021). Some authors speculated that
79 jSDMs applied to abundance data - instead of presence/absence data - are likely to provide a better
80 proxy for biotic interactions (Blanchet *et al.* 2020 ; Momal *et al.* 2020). Accordingly, jSDM have
81 progressively been extended and applied to abundance data (Hui 2016 ; Ovaskainen *et al.* 2017b ;
82 Chiquet *et al.* 2021 ; Popovic *et al.* 2022). Yet, specific challenges related to modelling abundance

83 data have only been recently explored in the context of species distribution modelling (Waldock
84 *et al.* 2022). To date, the predictive and the explanatory power of jSDM fitted to abundance data
85 remains largely untested compared to presence/absence data (Norberg *et al.* 2019 ; Wilkinson *et al.*
86 2020).

87 Second, regardless of the type of data considered (i.e. presence/absence or abundance), sev-
88 eral factors may limit or affect the interpretability and predictive ability of jSDM. For instance,
89 co-occurrence patterns estimated in jSDM are affected by unaccounted environmental variables im-
90 plying that jSDMs cannot fully separate the environmental and the biotic niche of species (Blanchet
91 *et al.* 2020 ; Poggiato *et al.* 2021). Beyond missing environmental predictors, one prerequisite
92 for improving biotic inference and thus jSDMs' predictions is to take into account other actors
93 (i.e. species) that could have an influence on the target community (e.g. competitors; Levine *et*
94 *al.* (2017)). However, because many ecological studies only focus on particular taxonomic groups
95 (Pollock *et al.* 2014 ; Häkkinen *et al.* 2018), hence disregarding non-target taxa, co-occurrence patterns
96 estimated from jSDMs are almost always skewed by missing ecological actors (Momal *et al.* 2021).
97 How this bias affects the predictive ability of jSDM remains untested.

98 Finally, similarly to SDMs, jSDMs can theoretically be extended to include additional sources
99 of information about modelled species (Niku *et al.* 2019 ; Ovaskainen *et al.* 2017b). For instance,
100 accounting for phylogenetic relationships between species (Ives & Helmus 2011) or for the link
101 between functional traits and environmental responses (Pollock *et al.* 2012) have been shown
102 to improve both the explanatory and the predictive powers of SDMs (Morales-Castilla *et al.* 2017
103 ; Vesik *et al.* 2021), which supports the hypothesis that similar species in terms of traits and/or
104 recent evolutionary history share similar environmental preferences. While similar effects related
105 to inclusion of species-specific information are expected in jSDMs (Ovaskainen *et al.* 2017b), the
106 relative influence of additional sources of information on their interpretability and predictive power
107 remains untested (Norberg *et al.* 2019 ; Wilkinson *et al.* 2019).

108 Overall, many practical questions remain concerning the application of jSDMs to ecological
109 community monitoring data in particular related to inclusion of additional sources of information
110 within the models. In this study, we aim to provide a comprehensive assessment of how jSDM
111 predictive and explanatory powers are affected by different sources of information. Specifically, by
112 comparing predictions obtained from a baseline model excluding additional sources of information
113 (i.e. a classical jSDM), we tested the effect of (1) including phylogeny alone and in combination with

114 trait data, (2) incorporating monitoring information related non-target species and (3) considering
115 abundance instead of presence/absence data. We hypothesised that all these sources of information
116 should improve jSDM predictive and explanatory powers, but did not assume a priori that a given
117 modelling strategy would lead to greater improvements in model performances.

118 **Materials & Methods**

119 We used the HMSC (Hierarchical Modeling of Species Communities) framework applied to the
120 long-term REBENT coastal monitoring dataset (rebent.ifremer.fr). In the following subsections,
121 we sequentially describe Fig. 1 : (1) the HMSC framework, (2) the data used in this study, (3) data
122 splitting between training and testing sets to assess the explanatory and predictive powers of
123 models, respectively, (4) the rationales for the suite of alternative models considered and, (5) the
124 performance metrics used to compare models.

125 **Hierarchical Modelling of Species Community (HMSC)**

126 “HMSC is a multivariate hierarchical generalised linear mixed model adjusted with Bayesian inference
127 rooted in assembly theory” (Ovaskainen & Abrego 2020). A HMSC model is composed of two
128 parts: one taking into account fixed effects and the other taking into account random effects. The
129 fixed part models the realised niche (i.e., the set of environmental conditions that is biotically
130 suitable and accessible to the species; Ovaskainen & Abrego (2020)) of each species (B matrix),
131 where each dimension of the niche is a covariate (e.g. temperature) included in the model (Fig. 1 ;
132 Ovaskainen & Abrego (2020)). Including trait data enables estimating of species-specific expected
133 niche value by accounting for trait-environment relationships, where species with similar traits are
134 expected to respond similarly along environmental gradients (Fig. 1 ; Ovaskainen *et al.* (2017b)
135 ; Ovaskainen & Abrego (2020)). It is well-established that phylogenetically-close species tend to
136 share similar trait values or niches (Wiens *et al.* 2010). Adding phylogenetic data to a HMSC model
137 already including traits is not necessarily redundant because it could capture residual ecological
138 information not included in the available trait data. This can theoretically improve species niche
139 estimates (Ovaskainen & Abrego 2020). Inclusion of such additional pieces of information can
140 moreover improve model fit for rare species by borrowing information on traits- (or phylogenetic-)
141 environment relationships estimated for common species that are similar in terms of traits (or
142 phylogenetic; Ovaskainen & Abrego (2020)). This property is a main advantage of hierarchical

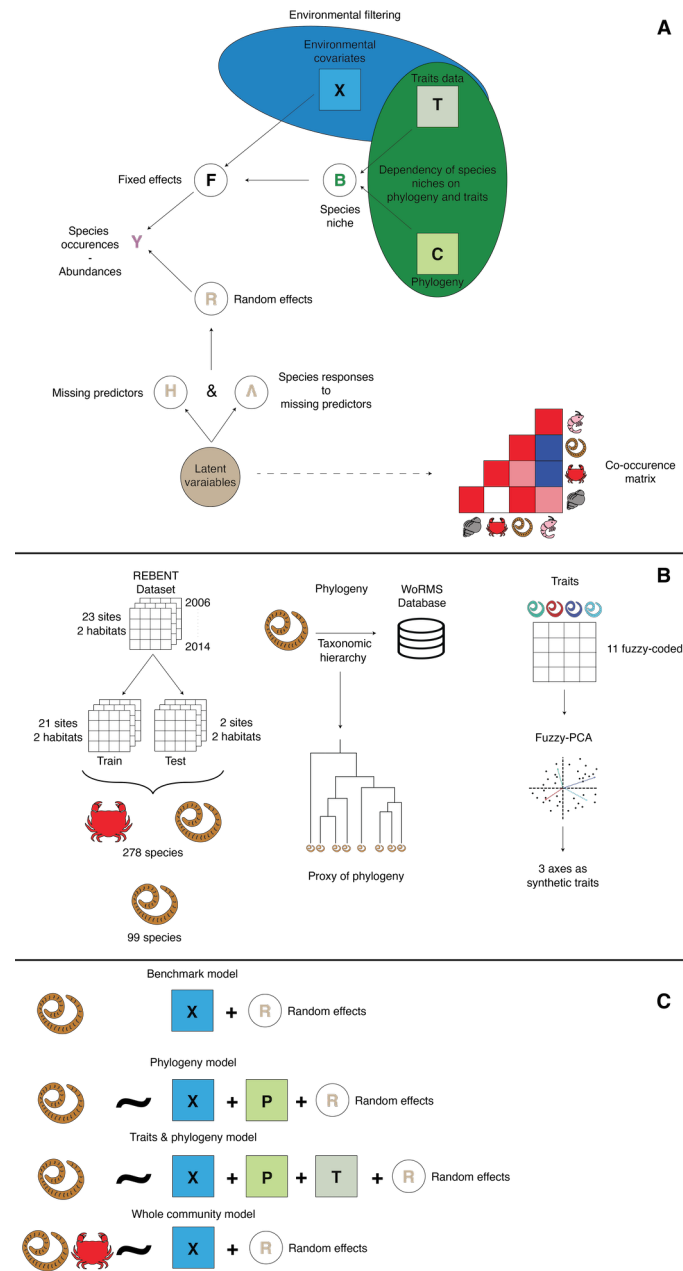


Figure 1. Workflow of the study. A. Structure of a Hierarchical Model of Species Community (HMSC) including environmental variables, phylogeny and species-specific functional traits. B. Data pre-processing: community data partitioning between train and test datasets, estimating of phylogenetic distance between species (using taxonomic classification) and dimension reduction of species-trait matrix using a fuzzy-PCA. C. Summary of the four alternative model structures fitted both on presence/absence and abundance data: the Benchmark, Phylogeny and Traits & Phylogeny models only considered polychaete species assemblage data, while the Whole Community model includes information related to additional species sampled alongside the target assemblage (i.e. non-target species that are not of direct interest but potentially interact with the target assemblage). Random effects accounting for the sampling year, site and habitat were included in all models.

143 models (Gelman *et al.* 2020).

144 The random part of HMSC relies on latent variables. Specifically, for each random effect, two ma-
145 trices of latent variables are estimated (Ovaskainen *et al.* 2017b ; Tikhonov *et al.* 2019 ; Ovaskainen
146 & Abrego 2020): the H matrix (called site loadings) contains the values of missing covariates not
147 included in the model (Fig. 1 ; Ovaskainen *et al.* (2017b) ; Ovaskainen & Abrego (2020)); while the Λ
148 matrix (called species loadings) corresponds to the response of the species to missing covariates
149 (Fig. 1 ; Ovaskainen *et al.* (2017b) ; Ovaskainen & Abrego (2020)). These covariates thus capture
150 residual variance, which can be due to various factors including missing environmental features or
151 the effect of biotic interactions (Ovaskainen *et al.* 2017a ; Ovaskainen *et al.* 2017b ; Ovaskainen &
152 Abrego 2020).

153 Datasets

154 Faunistic data

155 Faunistic data come from the REBENT programme (rebent.ifremer.fr), which is a station-based
156 ongoing monitoring network initiated in 2003 to detect, characterise and explain changes of coastal
157 benthic macrofauna across Brittany's coastline (Western France). Here, we focused on benthic
158 infaunal communities found in two soft-bottom habitats: intertidal bare sediments and intertidal
159 seagrass meadows (*Zostera marina*). Data from Boyé *et al.* (2019), covering 23 sites (Fig. S1)
160 monitored using the same protocol between 2006 and 2014, were used in this study. At each site,
161 sampling consists in the collection of three sediment cores of 0.03m² that are pooled together
162 and considered as a single sampling unit at each site. For each sampling event, individuals were
163 identified to the lowest taxonomic level possible (mostly species level; for simplicity we hereafter
164 use the term "species"). A detailed description of the sampling methodology is provided in (Boyé *et*
165 *al.* 2017 ; Boyé *et al.* 2019). Overall, across a total of 375 sampling units (i.e. unique combination of
166 years, sites and habitats), 861,997 individuals belonging to 821 species were collected and identified.

167 Functional traits and phylogeny data

168 We collated species-specific information related to functional traits and phylogeny for inclusion in
169 different models. These data were particularly well resolved for the polychaete community which
170 therefore constitutes the main object of inference. Polychaeta is a taxonomic group composed of
171 numerous species exhibiting diverse lifestyles (Jumars *et al.* 2015) that can be used to monitor the

172 health of benthic habitats (Giangrande *et al.* 2005). The polychaete traits data, which was available
173 for the 99 polychaete species present in the training set, includes 11 fuzzy-coded traits for a total
174 of 41 modalities (Boyé *et al.* 2019). Prior to jSDM fitting, the dimensionality of the trait matrix was
175 reduced using a fuzzy-PCA with the *fpca* function from the *ade4* R package (Thioulouse *et al.* 2018).
176 The first three axes, which account for 59% of the total variance of the trait matrix, were included in
177 the model as synthetic traits data (Fig. 1). The first axis distinguishes mobile predatory species from
178 sessile microphages; the second axis differentiates semelparous species from iteroparous species;
179 and, the third axis separates burrowers from tube-dwellers (Fig. S5).

180 In complement to the traits information available for the 99 polychaete species of interest, we
181 retrieved their taxonomic classification through the WoRMS database (www.marinespecies.org;
182 assessed in January 2020) and used this information as a proxy for phylogenetic relationships (Fig. 1
183 ; Ricotta *et al.* (2012) ; Ovaskainen & Abrego (2020)). Phylogenetic distances between polychaete
184 species were then estimated using the *ape* R package (Paradis & Schliep 2019).

185 Environmental data

186 Following Boyé *et al.* (2019), we selected seven environmental variables to characterise the ecologi-
187 cal niche of each species within the target community. These seven variables quantify different
188 components of coastal environmental variability including hydrology (sea water temperature, salin-
189 ity and current velocity), sedimentology (mud and organic matter content), substrate heterogeneity
190 (Trask index) and local wave exposure (fetch). For each of these seven variables, the first and second
191 degree polynomials were computed to account for non-linear responses.

192 Comparison of alternative model structures

193 The first model (benchmark model abbreviated as “Bench”) only relies on polychaete community
194 data and environmental covariates (Fig. 1). The second model (phylogenetic model abbreviated
195 as “Ph”) adds phylogenetic data to the Bench model (Fig. 1), which implies that rare species can
196 thus benefit from phylogenetic-environment relationships estimated for closely related species
197 (Ives & Helmus 2011). The third model (traits & phylogeny model abbreviated as “TrPh”) adds traits
198 data to the Ph model (Fig. 1), which means that rare species can benefit from traits-environment
199 relationships estimated for species presenting similar functional traits (whereas phylogeny can
200 capture ecological similarities between species, which are not captured by trait similarity; Pollock

201 *et al.* (2012)). Finally, the fourth model (whole community model abbreviated as “WhC”), adds
 202 information about the whole community (i.e. including non-polychaete species for a total of 278
 203 species) to the Bench model (only 99 polychaete; Fig. 1). This model does not include trait or
 204 phylogenetic data for the sake of computation time. Each of these four models were fitted twice,
 205 either using presence/absence or abundance data. All models include the same random effects
 206 (Fig. 1): a temporal random effect to account for variability across years, a spatial random effect
 207 to account for variability across sites and another spatial random effect to account for variability
 208 across habitats (bare vs seagrass).

209 **Model fitting and performance**

210 **Model fitting using Markov Chain Monte Carlo**

211 HMSC uses a Bayesian framework for model fitting where the posterior distribution is sampled using
 212 a MCMC algorithm. For each model we ran 15 chains, each with 30,000 iterations. The first 10,000
 213 iterations were discarded as burn-in while the remaining were thinned every 20 iterations yielding
 214 1,000 posterior samples per chain. Hence, in total, 15,000 posterior samples were recorded for
 215 each parameter. Model convergence for each model parameter was assessed using the potential
 216 scale reduction factor (Gelman & Rubin 1992).

217 **Assessing model performance and interpretability**

218 In order to independently assess models' predictive performance, we splitted the dataset into a train
 219 and a test set. The training dataset includes 180 sampling units defined as unique combinations
 220 of years (varies between six and nine depending on sites), sites (21) and two habitats (Fig. S1).
 221 From this dataset, that originally contained 519 species, we removed the species that occurred less
 222 than four times across the 180 observational units to avoid convergence issues and poor model
 223 inference, leading to the removal of 241 species. The remaining 278 species encompassed the 99
 224 polychaete species that made up the target community and the 142 non-target species that were
 225 included in the WhC model. The test dataset was composed of 35 sampling units related to all
 226 surveys over a 9-year period at two specific sites , where both habitats (i.e. seagrass and bare sand)
 227 occur. Beyond the presence of both habitats, these two sites were also chosen because they occur
 228 in environmental conditions that can be considered average at the scale of the region (thereby
 229 limiting extrapolation of the model; Boyé *et al.* (2017) ; Boyé *et al.* (2022) ; Toumi Chirine (n.d.))

230 while still harbouring different communities, representative of the known diversity gradient across
231 the region (Boyé *et al.* 2017 ; Toumi Chirine n.d.).

232 To investigate jSDM's performance, models were evaluated using a set of complementary metrics
233 to evaluate both their explanatory (predictions compared to observations of the train dataset) and
234 predictive (predictions compared to observations of the test dataset) powers (Wilkinson *et al.* 2020).
235 To assess models' performance, both overall (i.e., across all species) and at individual species level,
236 we used AUC and root mean squared errors (RMSE) for presence/absence and abundance models,
237 respectively. For the "whole community" model that most improved predictive power (see results),
238 we further explored species-specific gain in explanatory power by examining potential correlations
239 between (i) RMSE and the proportion of presences and (ii) RMSE and average abundance using the
240 Kendall rank correlation coefficient.

241 While the AUC and the RMSE provide estimates of model performance, either overall or for
242 individual species, these measures only partially capture model accuracy (or performance) at the
243 community scale. Hence, we also explored differences between observed and predicted community
244 composition (both for the train and test datasets) by decomposing the total beta diversity (using the
245 Sørensen index) into species turnover and nestedness using the *betapart.temp* function from the
246 *betapart* R package (Baselga 2010 ; Baselga *et al.* 2022). For abundance models, predictions were
247 transformed to presence/absence before computing beta diversity (i.e. all non-zero abundance
248 predictions were considered as presences). Thus, this framework assigns a total beta diversity of
249 zero to a model predicting the exact observed community, whereas a model predicting a completely
250 different community compared to observations is associated with a total beta diversity of one.
251 Moreover, using Baselga *et al.* (2022)'s framework, we decomposed beta diversity (i.e. predicted
252 error in community composition) according to two components: (1) turnover, if model correctly
253 estimates observed species richness but mispredicts species identity and (2) nestedness, if model
254 correctly predicts the identity of observed species but omits some.

255 To assess model interpretability, we calculated the proportion of explained variance attributed
256 either to environmental covariates (fixed effects) or to random effects. To evaluate the effect
257 of model structure on estimated species-environment relationships, we classified the shapes of
258 estimated response curves inferred from the different models according to nine categories that
259 characterise both their direction (decline, null or increase) and their acceleration (decelerated,
260 constant or accelerated) (Rigal *et al.* 2020). We then looked for differences between models in

terms of proportion of estimated response curves across each of these nine categories. Finally, to compare changes in random effects across models, we estimated differences between the Benchmark model and the best performing model in terms of residual co-occurrence patterns. We specifically quantified differences between models in both magnitude and sign of residual species-species correlations using the following index:

$$\text{Index} = |corr_{\text{best model}} - corr_{\text{benchmark}}| \times \text{sign}(corr_{\text{best model}} \times corr_{\text{benchmark}})$$

Results

Both MCMC convergence and effective sample size of the different jSDMs were satisfactory (see Appendix B).

Model Fit & Predictive power

Species level

Presence/absence models presented an excellent explanatory power as reflected by mean AUC estimates greater than 0.9 (Fig. S4). Conversely, their predictive power was rather low given a mean AUC estimate of ~0.65 (Fig. S4). For abundance models, mean RMSE computed on the training set ranged from 8.94 to 9.35 (Fig. S4). Their predictive power was heterogeneous with the whole community (WhC) model presenting the highest performance (mean RMSE = 5.83) followed by the benchmark model (Bench) (mean RMSE = 53.7), the phyloheny (Ph) model (mean RMSE = 63.7) and the traits & phylogeny model (mean RMSE = 95.3) (Fig. S4).

Relative to the benchmark model (Fig. 2), model explanatory power only slightly decreased for both TrPh (mean increase in RMSE +0.8% and decrease in AUC of -0.6%) and Ph models (mean increase in RMSE of +0.5% and decrease in AUC -0.6%). Explanatory power only slightly increased for the WhC models (mean decrease in RMSE of -3.6% and increase in AUC +0.3%). In terms of predictive power, performance mostly increased for the WhC abundance model with a mean decrease in RMSE of 26% relative to the benchmark model. This improvement concerned 62 species (mean decrease in RMSE of -49.3%; 10th and 90th deciles [-94.8% ; -9.7%] for these species) whereas 12 species presented a performance decrease (+36% RMSE; 10th and 90th deciles [10% ; 70.1%]). Only 32 and 36 species were improved for TrPH and Ph models, with mean decrease/increase in RMSE of 26.5% and 24.9% across all species, respectively.

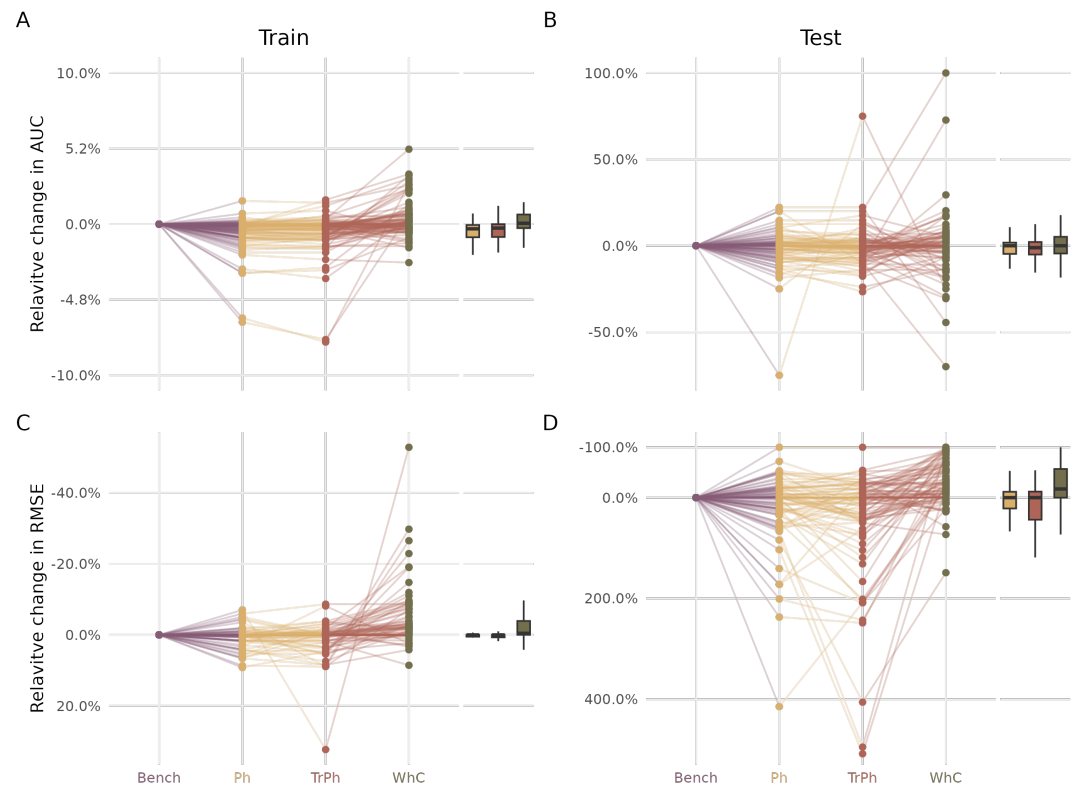


Figure 2. Relative change in explanatory (left column; panels A and C) and predictive (right column; panels B and D) power of different model structures (as colour-coded: purple for benchmark (Bench), yellow for phylogeny (Ph), red for traits & phylogeny (TrPh), and green for whole community (WhC) models). Relative changes (in %) are expressed relative to the benchmark fitted on presence/absence (top row; panels A and B) or abundance (bottom row; panels C and D) data. In all panels, points above the zero line (i.e. increase in AUC for panels A-B but decrease in RMSE for panels C-D) indicate performance gain.

Model explanatory performance increased for the most common (correlation between RMSE and proportion of presence: Kendall's $\tau = -0.28$, p-value $< 1e-5$) and abundant (correlation between RMSE and average species abundance: Kendall's $\tau = -0.29$, p-value $< 1e-4$).

Community level

On the training set, the median Sørensen dissimilarity index, which ranged from 0.36 to 0.38 across all models (both presence/absence and abundance), suggests that predicted communities are relatively similar to observed communities (Fig. S8 and Fig. S9). Errors were equally distributed between turnover and nestedness (Fig. S8 and Fig. S9). However, relative to observed communities in the test data set, abundance models predictions presented a median Sørensen dissimilarity of 0.65 while dissimilarity reached 0.72 for presence/absence models (Fig. S8 and Fig. S9). Greater

298 dissimilarity between predicted and observed communities in the test dataset relative to the training
299 dataset is a direct consequence of models' limited predictive power at the species level (see above
300 and Fig. S8 and Fig. S9). Note that proportion of nestedness errors is greater in the WhC model
301 than in other models, suggesting that this model tends to correctly predict the presence of a subset
302 of the observed species assemblage composition (Fig. S8 and Fig. S9).

303 **Variance partitioning**

304 The amount of variance explained by each model can be partitioned between environmental
305 covariates and random effects. For all models, environmental variables account for most (more
306 than $75\% \pm 18\%$, mean \pm s.d.) of the variance (Fig. S7). However, a larger part of variance is explained
307 by random effects in the WhC model compared to the Bench model (Fig. S7). Relative to the Bench
308 model, the median relative change in variance explained by random effects increased by 8.6% for
309 the Ph model, 19.9% for the TrPh model and 35.4% for the WhC model (Fig. 3). Similar results were
310 obtained for presence/absence models (Fig. 3).

311 **Species niche estimated**

312 For all abundance models, more than 60% of estimated response curves were flat, suggesting a lack
313 of ecologically meaningful species-environment relationships (Fig. 4). This proportion even reached
314 more than 80% for the WhC model. Almost no convex or concave responses curves were estimated
315 for abundance models. Meaningful species-environment relationships essentially included constant
316 or accelerated declines, which respectively represented $\sim 10\%$ and $\sim 15\%$ of estimated response
317 curves for the three models that do not include the whole community (i.e. Bench, TrPh and Ph)
318 (Fig. 4). For the WhC model, these percentages dropped to 4.62% and 9.24%, respectively (Fig. 4).
319 Similar results were obtained for presence/absence models (Fig. S10).

320 Both abundance- and presence/absence TrPh models (which include species functional traits)
321 reveal some meaningful trait-environment relationships between the first fuzzy-PCA axis and the
322 seven environmental predictors. This suggests that the occurrence of certain traits is likely favoured
323 (or hindered) under certain environmental conditions (Fig. S6). For instance, mobile predatory
324 species were more negatively affected by fetch than sessile suspensivores (Fig. S6). Moreover,
325 increase in organic matter concentration and decrease in current velocities were associated with
326 higher abundances of suspensivore populations.

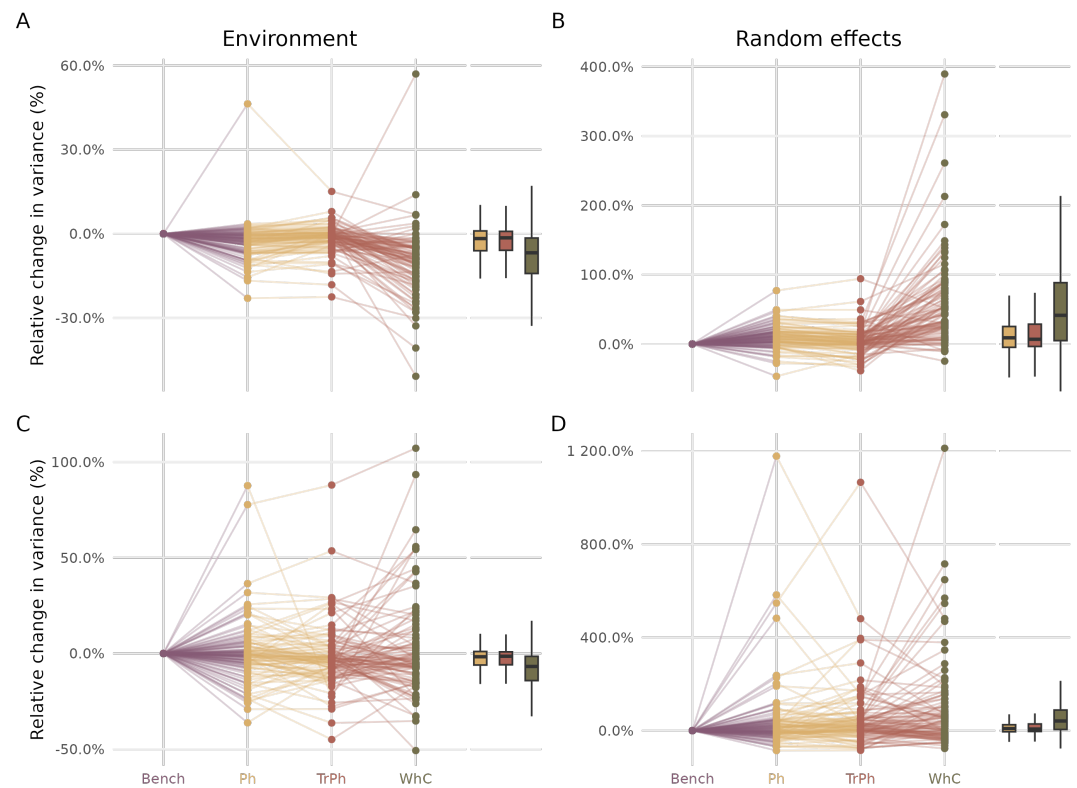


Figure 3. Relative change in variance explained by environmental predictors (left column) and by random effects (right column) according to alternative models as colour-coded (purple for Benchmark (Bench), yellow for phylogeny (Ph), red for traits & phylogeny (TrPh), and green for whole community (WhC) models). Percentage change is expressed relative to the benchmark model fitted with presence/absence (top panels) or abundance (bottom panels) data. In all panels, positive (negative) values point to an increase (decrease) in the proportion of variance explained by the focal model relative to the benchmark model.

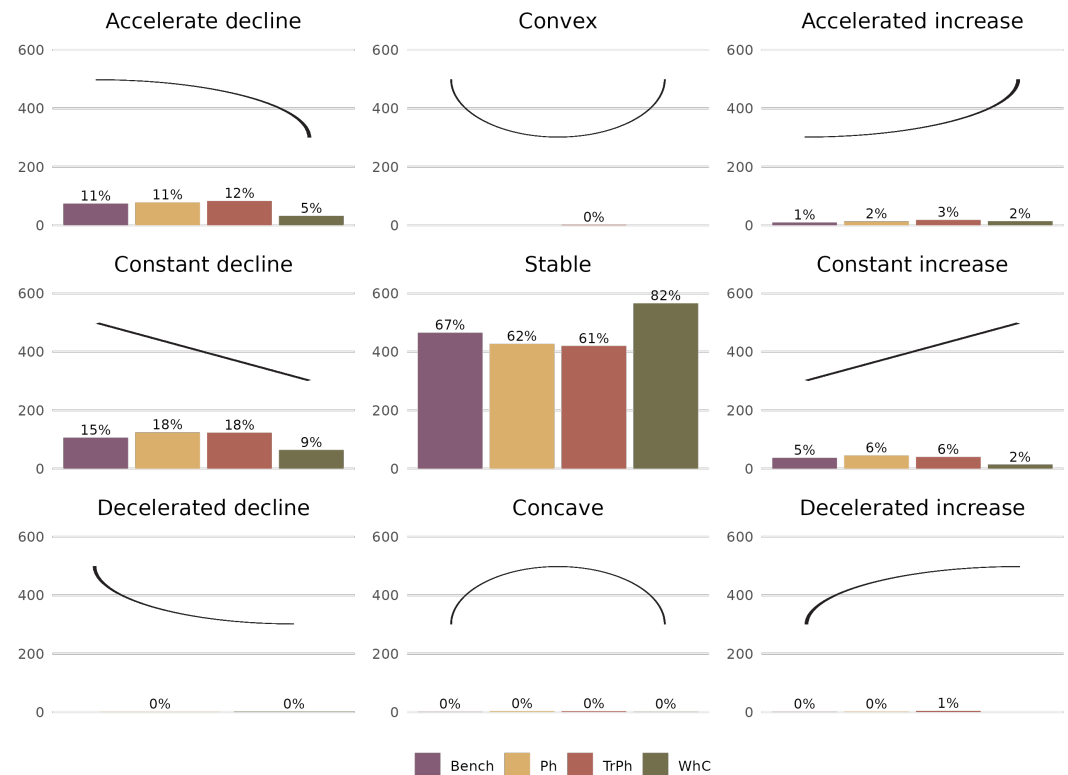


Figure 4. Number (y-axis) and proportion (indicated above individual bars) of response curves (i.e. one for each species-predictor combination) according to the nomenclature (nine shapes highlighted by the black curve in each panel) defined by Rigal *et al.* (2020) for different abundance model structures. See Fig. S10 for a similar representation for presence-absence models.

Exploring the residual correlation

Since all models included the same random effects, residual correlation matrices are comparable. Here, we compared residual correlations between the Bench model and the WhC model both when fitted to presence/absence or abundance data. We specifically considered the WhC model for this specific comparison, because of (1) its higher performances relative to alternative models and (2) the larger proportion of variance explained by random effects in this model relative to others (Fig. 3).

Residual correlations estimated by the WhC model were similar to those estimated by the Bench model, regardless of whether the models were fitted on presence/absence or abundance data (Fig. 5 and Fig. S11). Yet, agreement between models varied across the different random effects. For instance, when comparing residuals between Bench and WhC fitted on abundance data, correlation was low for random site effects ($R^2 = 0.66$), moderate for random habitat effects ($R^2 = 0.8$) and high for random year effects ($R^2 = 0.91$).

The dedicated comparison index helps qualify how pairwise species-species residual correlations change in sign and magnitude between the Bench and the WhC models. For abundance- or presence/absence models (Fig. 5 and Fig. S11), the index main modal distribution, which is centred on zero, suggests an overall agreement between residual correlations obtained from both models (Fig. 5 and Fig. S11). While the right part of the distribution highlights variation in the estimated magnitude of effect, on the left-hand part of the distribution (negative values) indicates a sign inconsistency in residual correlations between the two models. For abundance models, the Habitat, Site and Year random effects are respectively associated with 13.3%, 17.7% and 6% of sign inconsistencies in residual correlations between the Bench model and the WhC model. Similar results were obtained for presence/absence models.

Discussion

Case studies in community ecology typically rely on partial and heterogeneous observations (Pollock *et al.* 2020) but also on incomplete knowledge of target species ecological features (e.g. traits, phylogeny; Troudet *et al.* (2017)). In this paper we aimed to better understand how jSDM performance varies depending on the type of information included (i.e. phylogeny, traits or data on non-target species). While jSDMs have two main goals, i.e. explaining and predicting species dis-

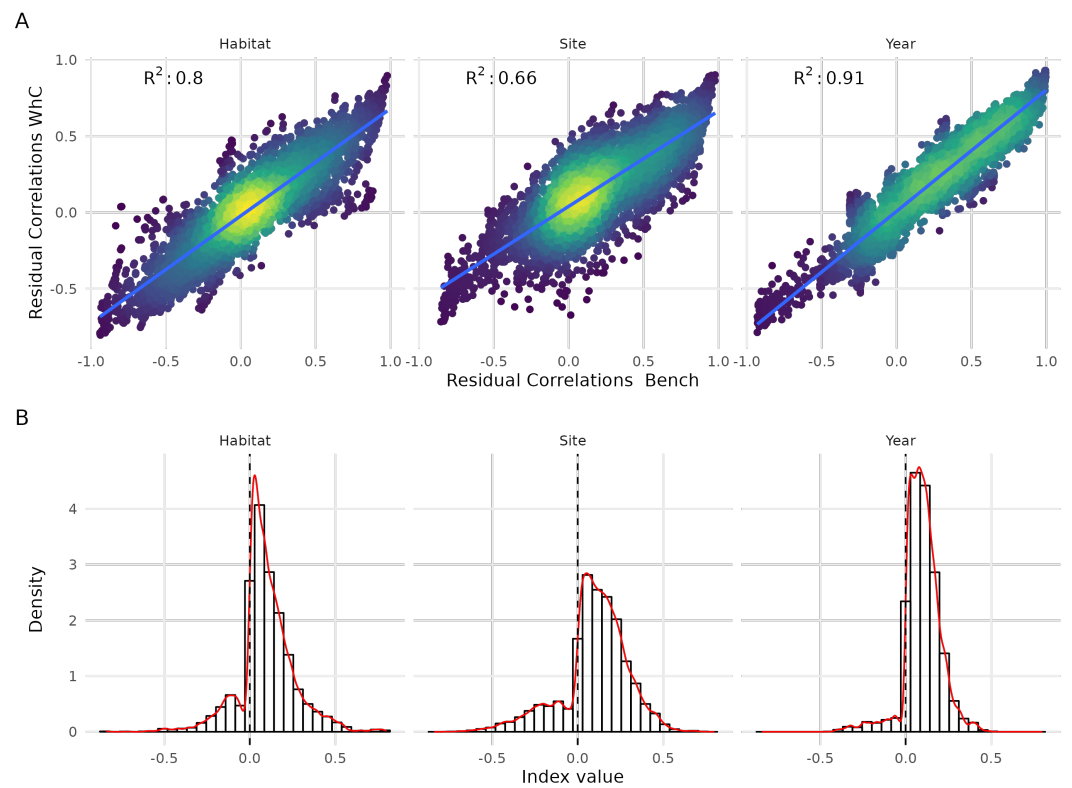


Figure 5. (A) Comparison of residual correlations associated with the three random effects estimated by the Whole Community Model (y-axis) and the Benchmark model (x-axis) fitted on abundance data. The colour scale highlights the density of points in each scatter plot. (B) Distribution of the index measuring change in sign (sign change left to the zero line, no change to the right) and magnitude (higher departure from the zero line indicate higher difference) between residual correlations estimated by the whole community model and the benchmark model adjusted with abundance data for the three random effects (Habitat, Site, Year).

356 tribution and community composition across space and/or time (Tredennick *et al.* 2021), to date,
357 they have mostly been tested with regards to their predictive power (Norberg *et al.* 2019), and
358 to some extent in terms of parameter estimates (Wilkinson *et al.* 2020), but only when fitted on
359 presence/absence data (Norberg *et al.* 2019 ; Wilkinson *et al.* 2020). Yet, jSDMs are increasingly
360 fitted on abundance data (e.g. Brimacombe *et al.* (2020)) and used for explanatory purposes (Abrego
361 *et al.* 2017; Häkkinen *et al.* 2018). Hence, there is currently a mismatch between the knowledge we
362 have regarding the performance of jSDMs and their application by ecologists. Here, we consolidate
363 the current understanding of jSDMs performance using complementary metrics and evaluation
364 methods. Overall, our results highlight that changes in model structure depending on the type of
365 information considered do impact many aspects of model performances (e.g. predictive power,
366 parameter estimates, estimated response curves, community composition). These changes can
367 have significant consequences on the interpretability and the conclusions drawn from these models,
368 especially for ecosystem management policies.

369 We found that jSDM's performance, in particular predictive power of abundance models, most
370 increased when including information related to the 179 non-target species sampled alongside
371 with the 99 polychaete species of interest. Given the hierarchical structure of HMSC (Poggiato *et al.*
372 *al.* 2021), inclusion of additional monitoring data related to other species likely improves model
373 performance by capturing a combination of relevant drivers that influence the target assemblage
374 but that are not explicitly considered in the model. For instance, inclusion of monitoring data for
375 other species can help describe the realised niche of the species of interest by capturing important
376 ecological processes related to environmental conditions (including trait-mediated responses)
377 or biotic interactions that are not well captured when considering only the target assemblage
378 (Ovaskainen *et al.* 2017a). In jSDMs, these unquantified ecological processes can be estimated using
379 latent variables from model residual correlation matrix. While this feature of jSDMs originally yielded
380 the potential to capture biotic interactions, it is now well-established that potential biotic signals
381 captured by jSDMs are largely confounded by other factors. These include missing environmental
382 variables (Dormann *et al.* 2018 ; Zurell *et al.* 2018 ; Blanchet *et al.* 2020), scale mismatch between
383 study organism responses and available environmental variables (Potter *et al.* 2013), coarse spatial
384 resolution of environmental variables (Zurell *et al.* 2018 ; König *et al.* 2021).

385 Importantly, while including non-target species improved predictive performance in our case
386 study, this does not mean that accounting for non-target species is always beneficial. These benefits

could indeed vary depending on robustness of non-target species monitoring data (e.g. detection issues), their role within the ecosystem (e.g. engineer species are likely more influential on local communities than rare transient species) or the process shaping the target assemblage (e.g. when the target assemblage is mostly under the influence of abiotic factors, then adding other species will have limited influence on model performance). Furthermore, how many non-target species should be included warrants further investigations. In the future, simulated datasets, overcoming limitations related to real world datasets (DiRenzo *et al.* 2022), could be used to more systematically assess how and to which extent jSDM performance is affected when increasing the number of non-target species and whether the functional or the trophic position of these species matters in this context. While species communities and assemblages are largely defined arbitrarily (Stroud *et al.* 2015), such sensitivity analysis could help delineate which ecological units to include (or not) to improve model performance for the species of management/conservation interests.

In practice, ecological studies often focus on a certain guild or taxonomic group (e.g. fish, birds) given data collection (consistent sampling methodology) or availability constraints (traits and/or phylogeny biased toward some taxonomic groups, Tyler *et al.* (2012) and usually centralised in taxonomic-centred repositories [e.g. FishBase; Froese & Pauly (2022)]), rather than for ecological reasons (e.g. all potential interactions well captured by the data at hand). In this study, dedicated focus on polychaetes was primarily guided by data availability (species-traits matrices available from Boyé *et al.* (2019) only included polychaetes) rather than for ecological reasons, although the fact that this taxonomic group is numerically dominant and highly diverse in terms of lifestyles and functional roles (Giangrande 1997 ; Jumars *et al.* 2015) was the reason that originally motivated trait-data collection.

jSDMs have already been used to model the distribution of a wide variety of species ranging from micro-organisms (Minard *et al.* 2019 ; Pichler & Hartig 2021) to megafauna (Rocha *et al.* 2017 ; Brimacombe *et al.* 2020) inhabiting many different ecosystems. Here, while we studied communities associated with two specific coastal habitats, i.e. seagrass and sand, that have original characteristics as they are located at the land-sea interface (Boyé *et al.* 2019), our case study reflects typical aspects of applied ecological research. These include issues related to data limitation and availability but also typical features of ecological communities (e.g. prevalence of rare and transient species; (Magurran & Henderson 2003 ; Snell Taylor *et al.* 2018)). Our results provide some insights on trait-environment relationships but these contributions of functional ecology in jSDMs are likely limited by trait data

quality and availability (Tyler *et al.* 2012 ; Juan *et al.* 2022). For instance, we found an interaction between trophic modalities (i.e. microphagous versus macrophagous diet) and fetch (Fig. S15), indicating that organisms that filter on small particles are less likely to occur in wave-exposed sites where high levels of sediment resuspension can block their filtering systems (Manning *et al.* 2014). Conversely macrophagous organisms are less impacted by fetch. Yet, most trait-environment relationships, and most species-environment relationships were flat suggesting that polychaete assemblages are driven by processes other than abiotic ones, including neutral processes (Boyé *et al.* 2019). However, the lack of contribution of other trait-environment relationships in our model could also be related to a mismatch between trait data, environmental data, and the ecological processes at play (Juan *et al.* 2022). For instance, the physical coastal environment is highly dynamic; a feature that is only partially characterised by our environmental variables that summarise average climatological conditions (but not extreme events or annual/seasonal variability). Likewise, the list of available fuzzy-coded traits only partially captures species capacity to adapt to frequent disturbances or environmental variability (Violle *et al.* 2012 ; Juan *et al.* 2022). Most ecological studies are likely to face similar trade-offs where the potential benefit of including traits within jSDMs is balanced out by the effort needed to collect relevant trait information when missing. In our case, while including traits does not improve model predictive power, it enhances our understanding of species responses along environmental gradients. Hence, if the goal is not prediction but inference (Tredennick *et al.* 2021), including traits and proxies of phylogeny can facilitate model interpretation, providing that explanatory power does not decrease (as in our case), and that additional model parameters do not make computation time impractical.

While guidelines have been developed to characterise the performance of jSDM fitted on presence-absence data (Wilkinson *et al.* 2020), it is only recently that the predictive power of abundance-based models has been explored (Waldock *et al.* 2022). Here, we used a set of complementary metrics to assess the performance of both presence-absence and abundance models at the species and community levels, the latter considering both alpha and beta diversity. We also transposed a method initially developed for time series (Rigal *et al.* 2020) to provide an innovative way of characterising the response curves of each species. Further, we bring together a set of approaches and propose a new index to characterise and compare residual correlation matrices. Overall, we provide a comprehensive framework for integrative assessment and comparison of alternative jSDM performance.

Overall, our results provide new insights into the most appropriate strategies for jSDM fitting, according to the objective of the modelling exercise (Troudet *et al.* 2017) and the data at hand. While the four considered models performed equally well in terms of explanatory power, adding extra information to traditional jSDM (i.e. that only consider abiotic predictors) can still prove useful in some cases. For instance, adding monitoring data for other non-target species can substantially increase model predictive power by modifying inferred species-environment relationships and residual correlation matrices. Similarly, adding traits or phylogeny can lead to better model interpretability. To confirm the generality of our findings, future studies should focus on ecosystems that display different characteristics such as stronger environmental filtering or competitive processes; or by resorting to simulated case studies (Zurell *et al.* 2010 ; DiRenzo *et al.* 2022).

Author Contributions

MPM conceived the original idea with inputs from CV, AB, MC. CV analysed the data and wrote the first draft of the manuscript. All authors provided critical comments on earlier version of the manuscript and gave final approval for publication.

Acknowledgment

The authors are very grateful to Marion Maguer and Vincet Le Garrec who conducted the field-work and laboratory analysis that made this study possible, as well as to all students, technicians, engineers and Ph.D. that have been involved in this long-term effort. This study was supported by the REBENT programme coordinated by Sandrine Derrien (MNHN) and its funding partners (Agence de l'eau Loire-Bretagne, Région Bretagne, DREAL Bretagne). The authors would also like to acknowledge the Pôle de Calcul et de Données Marines (PCDM) for providing DATARMOR storage and computational resources. <https://pcdm.ifremer.fr>. MPM is the recipient of an ANR early career grant ANR-21-CE02-0006.

References

- Abrego, N., Dunson, D., Halme, P., Salcedo, I. & Ovaskainen, O. (2017). [Wood-inhabiting fungi with tight associations with other species have declined as a response to forest management](#). *Oikos*, 126.
- Baselga, A. (2010). [Partitioning the turnover and nestedness components of beta diversity](#). *Global Ecology and Biogeography*, 19, 134–143.

- Baselga, A., Orme, D., Villeger, S., De Bortoli, J., Leprieur, F. & Logez, M. (2022). [betapart: Partitioning Beta Diversity into Turnover and Nestedness Components](#).
- Blanchet, F.G., Cazelles, K. & Gravel, D. (2020). [Co-occurrence is not evidence of ecological interactions](#). *Ecology Letters*.
- Boyé, A., Gauthier, O., Becheler, R., Le Garrec, V., Hily, C., Maguer, M., *et al.* (2022). [Drivers and limits of phenotypic responses in vulnerable seagrass populations: *Zostera marina* in the intertidal](#). *Journal of Ecology*, 110, 144–161.
- Boyé, A., Legendre, P., Grall, J. & Gauthier, O. (2017). [Constancy despite variability: Local and regional macrofaunal diversity in intertidal seagrass beds](#). *Journal of Sea Research*, 130, 107–122.
- Boyé, A., Thiébaud, Éric, Grall, J., Legendre, P., Broudin, C., Houbin, C., *et al.* (2019). [Trait-based approach to monitoring marine benthic data along 500 km of coastline](#). *Diversity and Distributions*, 25, 1879–1896.
- Brimacombe, C., Bodner, K. & Fortin, M.-J. (2020). [Inferred seasonal interaction rewiring of a freshwater stream fish network](#). *Ecography*, n/a.
- Brudvig, L.A. & Catano, C.P. (2022). [Prediction and uncertainty in restoration science](#). *Restoration Ecology*, n/a, e13380.
- Chesson, P. (2000). [Mechanisms of Maintenance of Species Diversity](#). *Annual Review of Ecology and Systematics*, 31, 343–366.
- Chiquet, J., Mariadassou, M. & Robin, S. (2021). [The Poisson-Lognormal Model as a Versatile Framework for the Joint Analysis of Species Abundances](#). *Frontiers in Ecology and Evolution*, 9.
- Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., *et al.* (2018). [Iterative near-term ecological forecasting: Needs, opportunities, and challenges](#). *Proceedings of the National Academy of Sciences*, 115, 1424–1432.
- DiRenzo, G.V., Hanks, E. & Miller, D.A.W. (2022). [A practical guide to understanding and validating complex models using data simulations](#). *Methods in Ecology and Evolution*, n/a.
- Dormann, C.F., Bobrowski, M., Dehling, D.M., Harris, D.J., Hartig, F., Lischke, H., *et al.* (2018). [Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions](#). *Global Ecology and Biogeography*, 27, 1004–1016.
- Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., *et al.* (2006). [Novel methods improve prediction of species' distributions from occurrence data](#). *Ecography*, 29, 129–151.
- Froese, R. & Pauly, D. (2022). [FishBase](#).

- 509 Gelman, A., Hill, J. & Vehtari, A. (2020). [Regression and Other Stories](#). Analytical Methods for Social
510 Research. Cambridge University Press.
- 511 Gelman, A. & Rubin, D.B. (1992). [Inference from Iterative Simulation Using Multiple Sequences](#).
512 *Statistical Science*, 7, 457–472.
- 513 Giangrande, A. (1997). [Polychaete reproductive patterns, life cycles and life histories: an overview](#).
514 In: *Oceanography And Marine Biology* (ed. A. D.Ansell, M.B., R. N.Gibson). CRC Press, pp. 310–411.
- 515 Giangrande, A., Licciano, M. & Musco, L. (2005). [Polychaetes as environmental indicators revisited](#).
516 *Marine Pollution Bulletin*, 50, 1153–1162.
- 517 Giannini, T.C., Chapman, D.S., Saraiva, A.M., Alves-dos-Santos, I. & Biesmeijer, J.C. (2013). [Improving](#)
518 [species distribution models using biotic interactions: a case study of parasites, pollinators and](#)
519 [plants](#). *Ecography*, 36, 649–656.
- 520 Godsoe, W., Franklin, J. & Blanchet, F.G. (2017). [Effects of biotic interactions on modeled species'](#)
521 [distribution can be masked by environmental gradients](#). *Ecology and Evolution*, 7, 654–664.
- 522 Häkkilä, M., Abrego, N., Ovaskainen, O. & Mönkkönen, M. (2018). [Habitat quality is more important](#)
523 [than matrix quality for bird communities in protected areas](#). *Ecology and Evolution*, 8, 4019–4030.
- 524 Holt, R.D. (2020). [Some thoughts about the challenge of inferring ecological interactions from spatial](#)
525 [data](#). *Biodiversity Informatics*, 15, 61–66.
- 526 Houlahan, J.E., McKinney, S.T., Anderson, T.M. & McGill, B.J. (2017). [The priority of prediction in](#)
527 [ecological understanding](#). *Oikos*, 126, 1–7.
- 528 Howard, C., Stephens, P.A., Pearce-Higgins, J.W., Gregory, R.D. & Willis, S.G. (2014). [Improving](#)
529 [species distribution models: the value of data on abundance](#). *Methods in Ecology and Evolution*,
530 5, 506–513.
- 531 Hui, F.K.C. (2016). [boral – Bayesian Ordination and Regression Analysis of Multivariate Abundance](#)
532 [Data in r](#). *Methods in Ecology and Evolution*, 7, 744–750.
- 533 IPBES. (2019). [Global assessment report on biodiversity and ecosystem services of the Intergovern-](#)
534 [mental Science-Policy Platform on Biodiversity and Ecosystem Services](#).
- 535 Ives, A.R. & Helmus, M.R. (2011). [Generalized linear mixed models for phylogenetic analyses of](#)
536 [community structure](#). *Ecological Monographs*, 81, 511–525.
- 537 Juan, S. de, Bremner, J., Hewitt, J., Törnroos, A., Mangano, M.C., Thrush, S., *et al.* (2022). [Biological](#)
538 [traits approaches in benthic marine ecology: Dead ends and new paths](#). *Ecology and Evolution*,
539 12, e9001.

- 540 Jumars, P.A., Dorgan, K.M. & Lindsay, S.M. (2015). [Diet of Worms Emended: An Update of Polychaete](#)
541 [Feeding Guilds](#). *Annual Review of Marine Science*, 7, 497–520.
- 542 Keil, P., Wiegand, T., Tóth, A.B., McGlinn, D.J. & Chase, J.M. (2021). [Measurement and analysis of](#)
543 [interspecific spatial associations as a facet of biodiversity](#). *Ecological Monographs*, n/a.
- 544 König, C., Wüest, R.O., Graham, C.H., Karger, D.N., Sattler, T., Zimmermann, N.E., *et al.* (2021). [Scale](#)
545 [dependency of joint species distribution models challenges interpretation of biotic interactions](#).
546 *Journal of Biogeography*, 48, 1541–1551.
- 547 Levine, J.M., Bascompte, J., Adler, P.B. & Allesina, S. (2017). [Beyond pairwise mechanisms of species](#)
548 [coexistence in complex communities](#). *Nature*, 546, 56–64.
- 549 Magurran, A.E. & Henderson, P.A. (2003). [Explaining the excess of rare species in natural species](#)
550 [abundance distributions](#). *Nature*, 422, 714–716.
- 551 Manning, L.M., Peterson, C.H. & Bishop, M.J. (2014). [Dominant macrobenthic populations experience](#)
552 [sustained impacts from annual disposal of fine sediments on sandy beaches](#). *Marine Ecology*
553 *Progress Series*, 508, 1–15.
- 554 Minard, G., Tikhonov, G., Ovaskainen, O. & Saastamoinen, M. (2019). [The microbiome of the *Melitaea*](#)
555 [cinxia butterfly shows marked variation but is only little explained by the traits of the butterfly](#)
556 [or its host plant](#). *Environmental Microbiology*, 21, 4253–4269.
- 557 Momal, R., Robin, S. & Ambroise, C. (2020). [Tree-based inference of species interaction networks](#)
558 [from abundance data](#). *Methods in Ecology and Evolution*, 11, 621–632.
- 559 Momal, R., Robin, S. & Ambroise, C. (2021). [Accounting for missing actors in interaction network](#)
560 [inference from abundance data](#). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*,
561 70, 1230–1258.
- 562 Morales-Castilla, I., Davies, T.J., Pearse, W.D. & Peres-Neto, P. (2017). [Combining phylogeny and](#)
563 [co-occurrence to improve single species distribution models](#). *Global Ecology and Biogeography*,
564 26, 740–752.
- 565 Niku, J., Hui, F.K.C., Taskinen, S. & Warton, D.I. (2019). [gllvm: Fast analysis of multivariate abundance](#)
566 [data with generalized linear latent variable models in R](#). *Methods in Ecology and Evolution*.
- 567 Norberg, A., Abrego, N., Blanchet, F.G., Adler, F.R., Anderson, B.J., Anttila, J., *et al.* (2019). [A compre-](#)
568 [hensive evaluation of predictive performance of 33 species distribution models at species and](#)
569 [community levels](#). *Ecological Monographs*, e01370.
- 570 Ovaskainen, O. & Abrego, N. (2020). *Joint Species Distribution Modelling: With Applications in R*. Ecology,

571 Biodiversity and Conservation. Cambridge University Press.

572 Ovaskainen, O., Tikhonov, G., Dunson, D., Grøtan, V., Engen, S., Sæther, B.-E., *et al.* (2017a). [How](#)
573 [are species interactions structured in species-rich communities? A new method for analysing](#)
574 [time-series data](#). *Proceedings of the Royal Society B: Biological Sciences*, 284, 20170768.

575 Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., *et al.* (2017b).
576 [How to make more out of community data? A conceptual framework and its implementation as](#)
577 [models and software](#). *Ecology Letters*, 20, 561–576.

578 Paradis, E. & Schliep, K. (2019). [ape 5.0: an environment for modern phylogenetics and evolutionary](#)
579 [analyses in R](#). *Bioinformatics*, 35, 526–528.

580 Pichler, M. & Hartig, F. (2021). [A new joint species distribution model for faster and more accurate](#)
581 [inference of species associations from big community data](#). *Methods in Ecology and Evolution*, 12,
582 2159–2173.

583 Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J.S. & Thuiller, W. (2021). [On the Interpre-](#)
584 [tations of Joint Modeling in Community Ecology](#). *Trends in Ecology & Evolution*.

585 Pollock, L.J., Morris, W.K. & Veski, P.A. (2012). [The role of functional traits in species distributions](#)
586 [revealed through a hierarchical model](#). *Ecography*, 35, 716–725.

587 Pollock, L.J., O'Connor, L.M.J., Mokany, K., Rosauer, D.F., Talluto, M.V. & Thuiller, W. (2020). [Protecting](#)
588 [Biodiversity \(in All Its Complexity\): New Models and Methods](#). *Trends in Ecology & Evolution*, 35,
589 1119–1128.

590 Pollock, L.J., Tingley, R., Morris, W.K., Golding, N., O'Hara, R.B., Parris, K.M., *et al.* (2014). [Understand-](#)
591 [ing co-occurrence by modelling species simultaneously with a Joint Species Distribution Model](#)
592 [\(JSDM\)](#). *Methods in Ecology and Evolution*, 5, 397–406.

593 Popovic, G.C., Hui, F.K.C. & Warton, D.I. (2022). [Fast model-based ordination with copulas](#). *Methods*
594 [in Ecology and Evolution](#), 13, 194–202.

595 Potter, K.A., Arthur Woods, H. & Pincebourde, S. (2013). [Microclimatic challenges in global change](#)
596 [biology](#). *Global Change Biology*, 19, 2932–2939.

597 Ricotta, C., Bacaro, G., Marignani, M., Godefroid, S. & Mazzoleni, S. (2012). [Computing diversity from](#)
598 [dated phylogenies and taxonomic hierarchies: does it make a difference to the conclusions?](#)
599 *Oecologia*, 170, 501–506.

600 Rigal, S., Devictor, V. & Dakos, V. (2020). [A method for classifying and comparing non-linear trajecto-](#)
601 [ries of ecological variables](#). *Ecological Indicators*, 112, 106113.

- 602 Rocha, R., Ovaskainen, O., López-Baucells, A., Farneda, F.Z., Ferreira, D.F., Bobrowiec, P.E.D., *et al.*
 603 (2017). [Design matters: An evaluation of the impact of small man-made forest clearings on](#)
 604 [tropical bats using a before-after-control-impact design](#). *Forest Ecology and Management*, 401,
 605 8–16.
- 606 Sander, E.L., Wootton, J.T. & Allesina, S. (2017). [Ecological Network Inference From Long-Term](#)
 607 [Presence-Absence Data](#). *Scientific Reports*, 7.
- 608 Snell Taylor, S.J., Evans, B.S., White, E.P. & Hurlbert, A.H. (2018). [The prevalence and impact of](#)
 609 [transient species in ecological communities](#). *Ecology*, 99, 1825–1835.
- 610 Staniczenko, P.P.A., Sivasubramaniam, P., Suttle, K.B. & Pearson, R.G. (2017). [Linking macroecology](#)
 611 [and community ecology: refining predictions of species distributions using biotic interaction](#)
 612 [networks](#). *Ecology Letters*, 20, 693–707.
- 613 Stroud, J.T., Bush, M.R., Ladd, M.C., Nowicki, R.J., Shantz, A.A. & Sweatman, J. (2015). [Is a community](#)
 614 [still a community? Reviewing definitions of key terms in community ecology](#). *Ecology and*
 615 *Evolution*, 5, 4757–4765.
- 616 Thioulouse, J., Dray, S., Dufour, A., Siberchicot, A., Jombart, T. & Pavoine, S. (2018). [Multivariate](#)
 617 [Analysis of Ecological Data with ade4](#). Springer.
- 618 Tikhonov, G., Opedal, O., Abrego, N., Lehtikoinen, A. & Ovaskainen, O. (2019). [Joint species distribution](#)
 619 [modelling with HMSC-R](#). *bioRxiv*.
- 620 Toumi Chirine, G.J., De Cáceres Miquel. (n.d.). Long-term coastal macrobenthic Community Trajec-
 621 tory Analysis reveals habitat-dependent stability patterns. *Ecography*.
- 622 Tredennick, A.T., Hooker, G., Ellner, S.P. & Adler, P.B. (2021). [A practical guide to selecting models](#)
 623 [for exploration, inference, and prediction in ecology](#). *Ecology*, 102, e03336.
- 624 Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. (2017). [Taxonomic bias in](#)
 625 [biodiversity data and societal preferences](#). *Scientific Reports*, 7.
- 626 Tyler, E.H.M., Somerfield, P.J., Berghe, E.V., Bremner, J., Jackson, E., Langmead, O., *et al.* (2012).
 627 [Extensive gaps and biases in our knowledge of a well-known fauna: implications for integrating](#)
 628 [biological traits into macroecology](#). *Global Ecology and Biogeography*, 21, 922–934.
- 629 Vesk, P.A., Morris, W.K., Neal, W.C., Mokany, K. & Pollock, L.J. (2021). [Transferability of trait-based](#)
 630 [species distribution models](#). *Ecography*, 44, 134–147.
- 631 Violle, C., Enquist, B.J., McGill, B.J., Jiang, L., Albert, C.H., Hulshof, C., *et al.* (2012). [The return of](#)
 632 [the variance: intraspecific variability in community ecology](#). *Trends in Ecology & Evolution*, 27,

244–252.

Waldock, C., Stuart-Smith, R.D., Albouy, C., Cheung, W.W.L., Edgar, G.J., Mouillot, D., *et al.* (2022). [A quantitative review of abundance-based species distribution models](#). *Ecography*, 2022.

Warton, D.I., Blanchet, F.G., O'Hara, R.B., Ovaskainen, O., Taskinen, S., Walker, S.C., *et al.* (2015). [So many variables: joint modeling in community ecology](#). *Trends in Ecology & Evolution*, 30, 766–779.

Whittaker, R.J., Willis, K.J. & Field, R. (2001). [Scale and species richness: towards a general, hierarchical theory of species diversity](#). *Journal of Biogeography*, 28, 453–470.

Wiens, J.J., Ackerly, D.D., Allen, A.P., Anacker, B.L., Buckley, L.B., Cornell, H.V., *et al.* (2010). Niche conservatism as an emerging principle in ecology and conservation biology. *Ecology letters*, 13, 1310–1324.

Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (2019). [A comparison of joint species distribution models for presence–absence data](#). *Methods in Ecology and Evolution*, 10, 198–211.

Wilkinson, D.P., Golding, N., Guillera-Arroita, G., Tingley, R. & McCarthy, M.A. (2020). [Defining and evaluating predictions of joint species distribution models](#). *Methods in Ecology and Evolution*, n/a.

Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., *et al.* (2010). [The virtual ecologist approach: simulating data and observers](#). *Oikos*, 119, 622–635.

Zurell, D., Pollock, L.J. & Thuiller, W. (2018). [Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments?](#) *Ecography*, 41, 1812–1819.