Is prediction of species richness from stacked species distribution models biased by habitat saturation?

Matthias Grenié, Cyrille Violle, François Munoz

Is prediction of species richness from Stacked Species Distribution Models biased by habitat saturation?

Matthias Grenié[1], Cyrille Violle[1], & François Munoz[2]

[1] CEFE, Univ. Montpellier, CNRS, EPHE, IRD, Univ. Paul Valéry Montpellier 3, Montpellier, France

[2] University Grenoble-Alpes, LECA, 2233 Rue de la Piscine, 38041 Grenoble Cedex 9, France

Author note

Correspondence concerning this article should be addressed to Matthias Grenié, . E-mail: matthias.grenie@ens-lyon.fr

Abstract

Several studies have proposed to predict Species Richness (SR) by combining the predictions of independent Species Distributions Models (SDMs) (the predict first-assemble later strategy). Alternative methods propose to combine outputs from SDMs differently, by either summing predicted presence probabilities at each location, or summing binary presence predictions after thresholding the probabilities. Species can occupy various proportions of their suitable habitats (i.e, have various levels of habitat saturation), which can cause discrepancy when predicting their presences through SDMs. Furthermore, these discrepancies can be increased when combining the predictions of individual SDMs to predict SR. In this article, we performed simulations of species distributions with varying habitat saturation (i.e., the amount of suitable habitat occupied by a species), and we compared observed richness with that predicted by the alternative approaches. We found that probability-based richness is not biased by the level of habitat saturation, while threshold-based richness over-predicts richness at low habitat saturation and under-predicts it as high habitat saturation. Probability-based richness should thus be used in priority when predicting species richness locally. Nonetheless, threshold-based richness represents species richness constrained by environmental filtering only and thus is a useful indicator of potential species richness when species fully saturate their habitats. Thus the systematic comparison of probability-based and threshold-based richness predictions can reveal the importance of habitat saturation and can thus help identify community assembly mechanisms at play.

Keywords: habitat saturation, species richness, stacked species distribution models, predicted presence probabilities, threshold-based presence prediction

Word count: 2909

Highlights

- Habitat saturation impacts predictions from Species Distribution Models (SDM)
- Habitat saturation biases stacked SDMs (S-SDMs) predictions
- Probability-based richness predicts local SR without bias whatever habitat saturation
- Comparing different S-SDMs predictions can shed light on community assembly processes

Introduction

Species Richness is an Essential Biodiversity Variable (EBV) (Pereira et al., 2013), which should be assessed, monitored and compared across space, time, and ecological contexts. Different models have been proposed for richness prediction in diverse ecological contexts and at large spatial scale (Dodson, 1992; Graham and Hijmans, 2006; O'Brien, 1998), with the perspective of identifying biodiversity hotspots (Mazel et al., 2014; Myers et al., 2000), targeting effective management practices (Chown et al., 2003), quantifying biodiversity changes (Newbold et al., 2015) and predicting ecosystem functioning (Cardinale et al., 2012).

Several methods can be used to predict richness depending on which ecological processes are at play. For example, Macro-Ecological Models (MEMs) directly predict richness at any location as a function of local environmental variables. These models consider the influence of environmental filtering and energy limits on richness (Hurlbert and Stegen, 2014). Because site-species data are first aggregated to estimate richness and then used to predict the variation with the environment, these approaches are called 'assemble first, predict later' (Ferrier and Guisan, 2006). Conversely, more and more global and local biodiversity databases include species

56 occurrences instead of local assemblage composition (GBIF, 2019; Sullivan et al., 2009; Tedesco
57 et al., 2017). An alternative approach has been to first model occurrences, independently for each
58 species, at any location using environmental variables through species distribution models
59 (SDMs) (Guisan and Thuiller, 2005; Guisan and Zimmermann, 2000), then to deduce potential
60 local richness by combining (=stacking) the predictions of individual SDMs (Calabrese et al.,
61 2014; D'Amen et al., 2015b; Gavish et al., 2017; Scherrer et al., 2018; Schmitt et al., 2017),
62 which is known as the 'predict first, assemble later' approach (Ferrier and Guisan, 2006). When
63 stacking SDMs, each SDM predicts occurrences for species independently using environmental
64 variables (Guisan and Zimmermann, 2000). Then, predictions of SDMs for different species are
65 summed to predict richness at assemblage-level. Stacked-SDMs (S-SDMs) predict observed
66 richness as well as or better than macro-ecological models (Dubuis et al., 2011; Guisan and
67 Rahbek, 2011), but there is still no consensus on the stacking method to be used so as to reliably
68 predict richness with S-SDMs (Scherrer et al., 2018).

69 Two main methods exist to stack SDMs (Dubuis et al., 2011; Pineda and Lobo, 2009;
70 Scherrer et al., 2018). Some authors suggested using thresholds to convert probabilities to binary
71 predictions (presence and absence) (Jim'enez-Valverde and Lobo, 2007; Liu et al., 2005). These
72 binary predictions are then summed to predict richness at local scale (hereafter threshold-based
73 richness). One of the main arguments for conversion of probabilities provided by SDMs to binary
74 predictions is that most of practical applications need binary maps (Jim'enez-Valverde and Lobo,
75 2007). A caveat of binary predictions is that they translate continuous responses of species along
76 environmental gradients into binary responses, which imply more abrupt shifts from presence to
77 absence between suitable and unsuitable conditions (Meynard and Kaplan, 2012). When
78 predicted probabilities are under the threshold, the model only predicts absences, while it only

predicts presences when predicted probabilities are above it. Close to the threshold value, a small change in predicted presence probability can change the binary prediction from absence to presence. Meynard et al. (2012) showed that presence predictions using thresholds fit observed presences only when species has a threshold-like response, while error increases when a species response is more gradual. The more species considered that have a gradual response along the environment, the greater the error when predicting richness. SDMs also directly provide continuous presence probabilities as outputs (Guisan and Thuiller, 2005), and threshold conversion to binary predictions adds a step compared to the direct sum of individual model predictions. Summing the probabilities of individual species model provides the mathematical expectation of the number of species locally present, assuming that species occurrences are independent (Calabrese et al., 2014; Violle et al., 2011), hereafter called probability-based richness.

A basic implicit assumption of SDMs is that only environmental conditions determine species occurrence, depending on a species fundamental niche (Guisan and Zimmermann, 2000). Additional processes should affect the realized occupancy patterns, such as dispersal limitation, competitive exclusion, local extinction dynamics (Pulliam, 2000). SDM predictions and thus richness predictions are likely to be biased by neglecting the contribution of processes shaping realized species distributions beyond their fundamental niche requirements (Václavík and Meentemeyer, 2012), thereby affecting SDM predictions and thus richness predictions. For instance, due to source-sink dynamics, some species can occupy less suitable sites, and thus be distributed outside the suitable habitat delimited based on presence probabilities predicted by SDMs. In addition, a species that is less often present across its suitable habitat would have a lower predicted presence probability than a species that is present in all its suitable habitats, even

though the predicted binary distribution of an SDM would be the same. We define habitat saturation of a species as a parameter that affects species occurrence probability based on environmental suitability. Here saturation is a species-level property and not an upper bound for richness in assemblages as proposed by Mateo et al. (2017). When species display low habitat saturation, their realized presence probabilities decrease, so that the predicted summed probability gets lower. On the contrary, the threshold-based presence prediction is not affected, by habitat saturation. Indeed, even if the determined species threshold changes with habitat saturation, the prediction will still be binary (presence or absence) (Meynard and Kaplan, 2012), thus we expect to observe increasing difference between threshold-based and observed richness with lower (or higher) habitat saturation.

Predicted presence probabilities partly reflect the ability of species to saturate their niche. Therefore, we expect probability-based richness to best predict actual richness. While we expect threshold-based richness to over-predict actual richness. Threshold-based richness rather represents a pool of species able to occur in given environmental conditions. To test these expectations we simulated virtual species with varying saturation and niche requirements (Hirzel et al., 2001; Meynard et al., 2019). We performed S-SDMs to predict richness given environmental conditions using both threshold- and probability-based richness and compared how the predictions were affected by habitat saturation. Probability-based richness followed observed richness whatever the habitat saturation, while threshold-based richness only matched observed richness when habitat saturation was 100%. Threshold-based richness only considered the environmental requirements of species, and could thus be used as the prediction of potential richness based solely on local environmental conditions. Potential richness could then be compared with other richness predictions that incorporate other ecological processes.

Material and Methods

Species assemblage simulations

Individual species simulation. We simulated a linear environmental gradient of 2000 values, from 1 to 2000. We then used the `virtualspecies` package version 1.4-2 (Leroy et al., 2016) to define 100 species independently, with quadratic environmental response $s_{i,k} = a \times \text{Env}_k^2 + b \times \text{Env}_k$, with $s_{i,k}$ the environmental suitability of species i in assemblage k and Env_k the environmental variable. a was drawn from a uniform distribution between -20 and -0.01. b was chosen as $b = -m * 2 * a$ where m was drawn from a uniform distribution between 1 and 2000 and represents the environment of maximum suitability. The suitability was then scaled between 0 and 1 by subtracting its minimum and dividing by the difference of its maximum and minimum. We used the function `generateRandomSp()` in `virtualspecies` to get suitability probabilities for each species and each environmental value (see Figure 1 left column).

Habitat saturation and predicted assemblages. We simulated species presences along the environmental gradient by performing binomial draws based on the presence probabilities. The presences probabilities $p_{i,k} = s_{i,k} \times \beta$ depend on (i) the suitability probabilities defined above, $s_{i,k}$ for species i and assemblage k, reflecting fundamental niche requirements, and (ii) an additional habitat saturation coefficient β representing the ability of species to occupy their suitable habitat (realized niche). When saturation is below 100%, the species tend to be less often present in suitable sites than species at 100% saturation (e.g., due to dispersal limitation or extinction). Species can also reach a saturation over 100% when they are present in less suitable conditions than according to their fundamental niche (e.g., through source-sink dynamics). We simulated 8 values of β: 10%, 40%, 70%, 100%, 120%, 150% and 170%. If the weighted

probability of presence was greater than one, we reduced it to a maximum of one. We thus simulated each species assemblage k for each value β.

Individual and Stacked Species Distribution Models

We performed Species Distribution Models (SDM) based on simulated species presences.

Modeling and Predicting Presences. We modeled the presence of each species using two predictors: the environmental value and the square of this value (see Figure 1 middle column) in Generalized Linear Models (GLM) of the binomial family:

$$\text{logit}(p_{i,k}) = \beta_0 \text{Env}_k + \beta_1 \text{Env}_k^2 \,, \tag{1}$$

with $p_{i,k}$ the presence of species i in assemblage k and Env_k its associated environmental variable. We thus estimated in each assemblage the probability of finding each species. For each species we determined the best threshold to get binary predictions by maximizing the True Skill Statistic (TSS) (Allouche et al., 2006). The TSS balances the proportion of presences correctly predicted and the proportion of absences correctly predicted.

Predicting Species Richness. We stacked SDM predictions in each assemblage to get a prediction of richness, with two approaches. We first summed the predicted presence probability for each species (probability-based richness, prediction **(A)** in Figure 1):

$$pred_{\text{rich,prob},k} = \sum_{i=1}^{S} p_i\,(k), \tag{2}$$

with $pred_{\text{rich,prob},k}$ the probability-based predicted richness in assemblage k, S the total number of species in the species pool, and $p_i(k)$ the predicted presence probability of species i in assemblage k. Using these probabilities we determine a species-specific threshold t_i using the True Skill Statistic (Allouche et al., 2006) that defines a binary function $1_i(k)$ to predict the presence of the species in each assemblage:

$$1_i(k) := \begin{cases} 1 & \text{if } p_i(k) \geq t_i \\ 0 & \text{if } p_i(k) < t_i \text{'} \end{cases} \quad (3)$$

with $p_i(k)$ the predicted presence probability of species i in assemblage k and t_i the species i threshold defined using TSS. We then compared the sum of predicted presence probabilities $pred_{\text{rich,prob},k}$ to the sum of predicted presences with species-specific threshold (threshold-based richness, prediction **(B)** in Figure 1):

$$pred_{\text{rich,thresh},k} = \sum_{i=1}^{S} 1_i\,(k), \quad (4)$$

with $pred_{\text{rich,thresh},k}$ the threshold-based predicted richness in assemblage k, S the total number of species in the species pool and $1_i(k)$ the indicator function defined as above.

We examined how predicted richness fitted observed richness across the whole environmental gradient, for different levels of habitat saturation. We quantified the deviation with Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N_k}\sum_{k=1}^{N_k}(pred_{\text{rich},k} - obs_{\text{rich},k})^2}, \quad (5)$$

176 with $pred_{\text{rich},k}$ the predicted richness of a given method in assemblage k, $obs_{\text{rich},k}$, the
177 observed richness in this assemblage, and N_k the total number of assemblages. We defined Bias
178 and Variance components:

$$\text{Bias} = \frac{1}{N_k}\sum_{k=1}^{N_k}(pred_{\text{rich},k} - obs_{\text{rich},k}) \quad (6)$$

$$\text{Variance} = \frac{1}{N_k}\sum_{k=1}^{N_k}(pred_{\text{rich},k} - \widehat{pred_{rich}})^2 \quad (7)$$

179 with $\widehat{pred_{rich}}$ the average predicted richness of a given method across all assemblages.

180 All analyses and SDMs were performed using R version 3.5.2 (R Core Team, 2019). A
181 version of the code used in this article is archived on Zenodo
182 (https://doi.org/10.5281/zenodo.3345742).

183 **Results**

184 Binary predictions (solid segments above and below the plot) showed few differences
185 whatever habitat saturation (Figure 2). There were the same from environment 1 to 273, then
186 between environment 467 and 1514, and for environments greater than 1720. In total binary
187 predictions were the same whatever habitat saturation for over 80% of the environmental values.

188 However, binary predictions changed abruptly from absences to presences and from presences to
189 absences for environment close to 500 and to 1500, respectively. On the contrary, the predicted
190 presence probabilities did vary with habitat saturation (solid curves in the center). The greater the
191 habitat saturation, the greater the maximum predicted probability. For example at 100% habitat
192 saturation, the maximum predicted probability was close to 0.95, while at 70% saturation it was
193 0.7.

194 When comparing observed richness to probability-based richness and threshold-based
195 richness (respectively green and purple points and curves on Figure 3), we observed differences
196 depending on habitat saturation. Across all habitat saturation levels, probability-based richness
197 showed consistently lower RMSE and variance than threshold-based richness (Figure 4). For
198 habitat saturation below 100%, threshold-based richness was greater than observed richness,
199 while probability-based richness followed observed richness. Observed richness against
200 probability-based richness followed the identity line closely with a slope not different from one
201 (all $p > 0.5$, H_0 being that the slope is not different from one) and an intercept not different from
202 zero, related to zero bias at all habitat saturation levels (Figure 4 middle). The relationship
203 between observed richness and threshold-based richness was not linear and did not follow the
204 identity line whatever habitat saturation. Probability-based richness showed similar RMSE at all
205 habitat saturation levels, while threshold-based richness reached its minimum RMSE when
206 habitat saturation was 80%. When species under-saturated their habitats ($\beta < 100\%$), probability-
207 based richness followed closely observed richness while threshold-richness almost always over-
208 predicted richness. Threshold-based richness lowest RMSE at 80% habitat saturation can be
209 explained by a balance between slight under-prediction when richness was smaller than 75 and
210 over-prediction when richness was greater than 75 (Figure 3). When habitat saturation reached

100%, both types of predictions were close to observed richness (Figure 3). At this habitat saturation, threshold-based richness showed slight over-prediction in richer communities (predicted richness around 90 species for sites containing 80 species) and slight under-prediction in poorer sites (predicted richness of around 30 for sites containing 45 species), and an average under-prediction (negative bias). At this habitat saturation, the RMSE of both methods was close to the one at 80% habitat saturation, but the variance in prediction increased for probability-based richness. When species over-saturated their habitats ($\beta > 100\%$), threshold-based richness strongly under-predicted richness in poorer communities (negative bias) while probability-based richness showed no bias on average (Figure 4 middle). For example at 150% habitat saturation, for sites with observed richness around 75, threshold-based richness was around 30 while probability-based richness was 75.

Discussion

We designed a virtual experiment of species occurrences along an environmental gradient and performed binomial GLM-based species distribution modeling on these data. The binary threshold-based presence prediction represented the potential habitat of each species based on its fundamental niche (Guisan and Zimmermann, 2000), whatever its actual habitat saturation. On the contrary, the range and average values of predicted presence probabilities depended on habitat saturation, for a given fundamental niche. When summing the individual species predictions, the summed presence probabilities well fitted actual richness, as expected, while habitat saturation strongly affected the threshold-based richness. We thus recommend summing stacked-SDMs probabilities to predict richness. Still, threshold-based richness can also be a

232 useful predictor of potential richness, as species threshold-based binary predictions can be used
233 as a reference species pool for hypothesis testing and modeling of biodiversity dynamics.

234 In our simulations, probability-based richness on average followed observed richness
235 whatever habitat saturation. This is in line with the fact that probability-based richness should
236 provide the mathematical expectation of richness at a given site (Calabrese et al., 2014). Our
237 results also showed that probability-based richness had a consistently lower RMSE than
238 threshold-based richness, mostly because of its absence of bias. However, both methods had
239 higher variance with higher habitat saturation, as a consequence of a mean-variance relationships.
240 Thus at high habitat saturation, both methods predict an unreliable richness.

241 Much emphasis has been put in species distribution modeling on providing binary
242 occurrence prediction. Methods to define thresholds for reliable occurrence prediction have been
243 extensively debated and alternative options have been proposed (Allouche et al., 2006; Freeman
244 and Moisen, 2008; Liu et al., 2005, 2013). However, such a prediction does not grasp the
245 inherently gradual response of species to environmental gradients (Hutchinson, 1957; Meynard
246 and Kaplan, 2012), and tends to generate an artificial dichotomy. This "binarization" has two
247 major caveats. First, it does not acknowledge the gradual variation of performance along the
248 gradient, which increases under-prediction below the threshold and over-prediction above the
249 threshold. Furthermore, the closer to the threshold the higher the prediction bias: just over/below
250 the threshold, there is a greater chance to find a species present/absent than further away from the
251 threshold. Second, it predicts only presences above the threshold and only absences below the
252 threshold, which does not acknowledge the influence of habitat saturation irrespective of habitat
253 suitability. In other words, threshold-based richness will always estimate richness as if species
254 habitat saturation was 100%. Because threshold-based richness over-predicts richness for habitat

255 saturation under 100% (or under-predicts when habitat saturation is over 100%), its accuracy
256 regarding the prediction of species turnover may be low (D'Amen et al., 2015b; Dubuis et al.,
257 2011). At coarser and larger scales, because niche preferences dominate the distribution of
258 species (Pearson and Dawson, 2003), we expect a more deterministic response to the
259 environment in a threshold-like fashion (Guisan and Thuiller, 2005). Species response to
260 environmental gradients is thus highly scale-dependent, specific at local and fine scales and
261 threshold-like at large and coarse scales (Meynard and Kaplan, 2013). The assumption that
262 species distribution at large and coarse scales is in a threshold fashion (Guisan and Thuiller,
263 2005) has been difficult to prove (Boucher-Lalonde et al., 2014, 2012). Instead in birds,
264 mammals and North American trees, a Gaussian distribution best explained the occurrence-
265 environment relationship for most species (Boucher-Lalonde et al., 2014, 2012), while the
266 threshold model was selected only 5% of the time. Only a fraction of species responds to broad
267 environmental gradients in a binary way. Meynard et al. (2013) also argued that threshold
268 response of species observed in many datasets could be the results of data aggregation over
269 various spatial and temporal scales.

270 We defined habitat saturation as a coefficient (β) that affects environmental suitability of
271 species: it increases ($\beta > 1$) or decreases ($\beta < 1$) habitat suitability. It has been shown in diverse
272 taxa that most species do not saturate their habitat: they occupy less habitat than their potential
273 habitat (Boucher-Lalonde et al., 2012; Munguia et al., 2008; Svenning and Skov, 2004). Several
274 mechanisms can explain why a species under-saturates its habitat. For example, dispersal
275 limitation due to slow recolonization of European trees from glacial refugia has led to habitat
276 under-saturation (Svenning and Skov, 2004). Biotic interactions are often cited as an additional
277 factor explaining habitat under-saturation (Svenning and Skov, 2004), as species close in traits

can experience limiting similarity and competitively exclude one another. On the contrary, positive biotic interactions as well as source-sink dynamics can cause habitat over-saturation (Eriksson, 1996; Pulliam, 2000; Pulliam and Danielson, 1991). Positive interactions such as facilitation make facilitated species occupy less suitable habitat thanks to the presence of other species (Bertness and Callaway, 1994; Stachowicz, 2001). Source-sink theory explains how a species can be present in unsuitable habitat (sink) by continuously immigrating from a suitable habitat (source) (Pulliam and Danielson, 1991). Here we considered a single habitat saturation coefficient (β) used for all species across all assemblages. This coefficient does not take into account the variability of habitat saturation that may exist between species, where some species saturate more their habitat then others. Furthermore, the habitat saturation coefficient cannot take into account biotic interactions as it is not conditional to the presence of other species; nor that we expect source-sink dynamics to occur only close to the sources, which should lead to a context-dependent habitat saturation. Habitat saturation is also influenced by the extent to which it is measured. In very small areas (e.g., a single quadrat), species tend to fully saturate their suitable habitat, because they occupy the only micro-habitat available for them. For larger areas (e.g., several plots), the occurrence of species should be more stochastic due to dispersal limitation, limiting similarity and biotic interactions as stated above. For even larger areas (e.g., regional, continental or global), habitat saturation should increase again with the dominance of deterministic processes that influence occurrence. As such, we could use a species habitat saturation profile at different areas whose variation would show the change in main assembly processes. Further research is needed regarding habitat (un-)saturation to understand its causes and consequences.

A recent study mentions a different but related concept of saturation (Mateo et al., 2017). Mateo et al. (2017) defined saturation as "environmental constraints [that] limit the number of species that can coexist in a community". Here, we defined habitat saturation as a species-level pattern: it represents the proportion of suitable habitat that a species occupies, it is a species-level property not a community-level property. Community saturation, i.e. saturation *sensu* Mateo et al. (2017), depends on habitat saturation of species. If species have a limited habitat saturation, it imposes an upper bound to species richness. In our model, there are no strict limits on species richness, but on each species' capacity to saturate its habitat. The neutral theory imposes a limit on the number of individuals in any community (Hubbell, 2001), it is a subset of the invidividuals/species present in the species pool that encompasses a larger area. Changes in number of individuals per community, species regional abundances and/or immigration probability m from the species pool can cause changes in species habitat saturation. By choosing these three parameters, we can obtain a stable richness that can be interpreted as an upper bound, as if saturation was community-level process. However, in this case there is no direct bound on species richness. Richness results from the dynamic extinction-colonization equilibrium and fluctuates over time, it is not a property of the community *per se*. Mateo et al. (2017) focused mostly on S-SDMs that can be be constrained with an explicit constraint on richness. As stated above, species richness is unlikely to be directly constrained and thus modeling explicitly a richness constraint may not underline the true community assembly mechanisms that affect community composition.

Our model, while an interesting basis to test assumptions regarding SDM stacking, represents an ecologically idealistic situation that uses virtual species. We used a single linear environmental gradient, which is an over-simplification of environmental gradients. Indeed,

323 species occurrences are jointly affected by multi-dimensional environmental gradients, which can
324 be non-linear and lead to observed trait syndromes (Laughlin and Messier, 2015). In our
325 simulations, all species have a single trait with a single optimum, however with multi-
326 dimensional environmental gradients we could also expect multi-dimensional optima (Oksanen
327 and Minchin, 2002). Our simulations do not consider biotic interactions, as we simulated the
328 presence of species independently, while, as stated above, biotic interactions can strongly
329 influence species habitat saturation (Pulliam, 2000). We used a species pool containing species
330 with optima on the whole range of the environmental gradient. However, the distribution in
331 species optima among the species pool can be asymmetrically distributed, which in turn can
332 affect local community assembly dynamics (Patrick and Brown, 2018). Furthermore, because of
333 the way the species were simulated, most species niche breadth covered around one third of the
334 range of the environment. While real communities contain a mix of species with narrow and wide
335 niches, with many of species having narrow niches and a few having wide niches (Brown, 1984).
336 Thus, we could determine a ratio between species with wide and narrow environmental niches,
337 based on observed communities, and simulate virtual communities accordingly. We also assumed
338 that species' suitabilities had a quadratic response to the environment, while more complex
339 relationships exist (Oksanen and Minchin, 2002) and could be used in our model. Our simulation
340 setup can thus be made more complex for more investigations on factors that may influence S-
341 SDMs richness predictions. Still, our simplified model can help gain insights about S-SDMs.

342 Depending on the scales considered, we can expect different shapes of species occurrence-
343 environment relationships. At local scale, we expect many stochastic processes (e.g.,
344 demographic stochasticity, competitive exclusion, biotic interactions, microclimatic variations,
345 etc.) to be at play and drive community assembly (Chase and Myers, 2011). Dominance of

346 stochastic processes leads to blurred response to environmental variables, because species
347 occurrence is then not only determined by environmental variables. Predicted presence
348 probability can account for these processes, because they predict the parameter that governs the
349 stochastic process leading to species occurrence such as a binomial trial (Pottier et al., 2013).
350 Indeed, probability-based richness have been shown to estimate the richness of local assemblages
351 well (Calabrese et al., 2014; D'Amen et al., 2015a, 2015b; Guisan and Rahbek, 2011; Pellissier et
352 al., 2013).

353 Threshold-based richness can be thought as the potential richness expected considering
354 only abiotic deterministic processes. It can be a useful baseline to compare to models that
355 consider a broader set of processes (Pouteau et al., 2019; Violle et al., 2011). Threshold-based
356 richness defines a reference pool against which null models or hierarchical analyses can be
357 performed. It can be considered as an additional method to define species pool (Carstensen et al.,
358 2013; Lessard et al., 2012). Indeed, threshold-based richness would represent a species pool
359 (Lessard et al., 2015) that considers only the response to environmental filtering for a large area.
360 Without explicitly considering functional traits, threshold-based richness can also represent a
361 functional species pool as species traits are filtered by the environment (de Bello et al., 2012).
362 Threshold-based richness is nested in a hierarchy of models similar to the hierarchy of scales and
363 processes that shape community assembly (Keil et al., 2013; Mackey and Lindenmayer, 2001;
364 Mertes and Jetz, 2018; Meyer, 2007; Pearson and Dawson, 2003). Threshold-based predictions,
365 because they consider environmental filtering only, are representative of coarse and large scales
366 in this hierarchy of models. In the SESAM framework (D'Amen et al., 2015a; Guisan and
367 Rahbek, 2011), threshold richness is a reference richness before applying a cutoff in species
368 presences to account for local variations. The use of threshold-based predictions can thus be

compared to more mechanistic models, to know to what extent observed communities are mostly shaped by environmental filtering. Other models, to which they can be compared, can incorporate other important community assembly factors such as dispersal limitation, limiting similarity or biotic interactions (Chase and Myers, 2011; Munoz et al., 2017; Pouteau et al., 2019; van der Plas et al., 2015). In summary, using both threshold-based richness and probability-based richness in succession —first threshold-based richness as a pure environmental prediction then compare it to probability-based richness— can shed light on community assembly processes. When both agree, environmental filtering dominates community assembly. If not, habitat saturation can strongly change threshold-based richness and/or other processes may affect community assembly. Threshold-based richness and probability-based can further be compared to other predictions using process-based models that consider additional processes. Essential Biodiversity Variables can be measured using multiple methods (Pereira et al., 2017), and there is no clear recommendation on which method should be prioritized to predict EBVs. Probability-based species richness could be used as a reliable method to predict taxonomic diversity in the EBV framework, while threshold-based richness can be a useful tool to assess community assembly processes (Pouteau et al., 2019).

Acknowledgments

We would like to thank Pierre Denelle and Christine Meynard for helpful discussions. MG was supported by the ENS de Lyon. This study was supported by the European Research Council (ERC) Starting Grant Project 'ecophysiological and biophysical constraints on domestication in crop plants' (grant ERC-StG-2014-639706-CONSTRAINTS) and by the French Foundation for

Research on Biodiversity (FRB; <www.fondationbiodiversite.fr>) in the context of the CESAB project 'causes and consequences of functional rarity from local to global scales' (FREE).

References

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS): Assessing the accuracy of distribution models. Journal of Applied Ecology 43, 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x

Bertness, M.D., Callaway, R., 1994. Positive interactions in communities. Trends in Ecology & Evolution 9, 191–193. https://doi.org/10.1016/0169-5347(94)90088-4

Boucher-Lalonde, V., Morin, A., Currie, D.J., 2014. A consistent occupancyClimate relationship across birds and mammals of the Americas. Oikos 123, 1029–1036. https://doi.org/10.1111/oik.01277

Boucher-Lalonde, V., Morin, A., Currie, D.J., 2012. How are tree species distributed in climatic space? A simple and general pattern. Global Ecology and Biogeography 21, 1157–1166. https://doi.org/10.1111/j.1466-8238.2012.00764.x

Brown, J.H., 1984. On the Relationship between Abundance and Distribution of Species. The American Naturalist 124, 255–279. https://doi.org/10.1086/284267

Calabrese, J.M., Certain, G., Kraan, C., Dormann, C.F., 2014. Stacking species distribution models and adjusting bias by linking them to macroecological models: Stacking species distribution models. Global Ecology and Biogeography 23, 99–112. https://doi.org/10.1111/geb.12102

411 Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A.,
412 Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace,
413 J.B., Larigauderie, A., Srivastava, D.S., Naeem, S., 2012. Biodiversity loss and its impact
414 on humanity. Nature 486, 59–67. https://doi.org/10.1038/nature11148

415 Carstensen, D.W., Lessard, J.-P., Holt, B.G., Krabbe Borregaard, M., Rahbek, C., 2013.
416 Introducing the biogeographic species pool. Ecography 36, 1310–1318.
417 https://doi.org/10.1111/j.1600-0587.2013.00329.x

418 Chase, J.M., Myers, J.A., 2011. Disentangling the importance of ecological niches from
419 stochastic processes across scales. Philosophical Transactions of the Royal Society B:
420 Biological Sciences 366, 2351–2363. https://doi.org/10.1098/rstb.2011.0063

421 Chown, S.L., van Rensburg, B.J., Gaston, K.J., Rodrigues, A.S.L., van Jaarsveld, A.S., 2003.
422 Energy, Species Richness, and Human Population Size: Conservation Implications at a
423 National Scale. Ecological Applications 13, 1233–1241. https://doi.org/10.1890/02-5105

424 D'Amen, M., Dubuis, A., Fernandes, R.F., Pottier, J., Pellissier, L., Guisan, A., 2015a. Using
425 species richness and functional traits predictions to constrain assemblage predictions from
426 stacked species distribution models. Journal of Biogeography 42, 1255–1266.
427 https://doi.org/10.1111/jbi.12485

428 D'Amen, M., Pradervand, J.-N., Guisan, A., 2015b. Predicting richness and composition in
429 mountain insect communities at high resolution: A new test of the SESAM framework.
430 Global Ecology and Biogeography 24, 1443–1453. https://doi.org/10.1111/geb.12357

de Bello, F., Price, J.N., Münkemüller, T., Liira, J., Zobel, M., Thuiller, W., Gerhold, P., Götzenberger, L., Lavergne, S., Lepš, J., Zobel, K., Pärtel, M., 2012. Functional species pool framework to test for biotic effects on community assembly. Ecology 93, 2263–2273. https://doi.org/10.1890/11-1394.1

Dodson, S., 1992. Predicting crustacean zooplankton species richness. Limnology and Oceanography 37, 848–856. https://doi.org/10.4319/lo.1992.37.4.0848

Dubuis, A., Pottier, J., Rion, V., Pellissier, L., Theurillat, J.-P., Guisan, A., 2011. Predicting spatial patterns of plant species richness: A comparison of direct macroecological and species stacking modelling approaches. Diversity and Distributions 17, 1122–1131. https://doi.org/10.1111/j.1472-4642.2011.00792.x

Eriksson, O., 1996. Regional Dynamics of Plants: A Review of Evidence for Remnant, Source-Sink and Metapopulations. Oikos 77, 248–258. https://doi.org/10.2307/3546063

Ferrier, S., Guisan, A., 2006. Spatial modelling of biodiversity at the community level. Journal of Applied Ecology 43, 393–404. https://doi.org/10.1111/j.1365-2664.2006.01149.x

Freeman, E.A., Moisen, G.G., 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecological Modelling 217, 48–58. https://doi.org/10.1016/j.ecolmodel.2008.05.015

Gavish, Y., Marsh, C.J., Kuemmerlen, M., Stoll, S., Haase, P., Kunin, W.E., 2017. Accounting for biotic interactions through alpha-diversity constraints in stacked species distribution models. Methods in Ecology and Evolution 8, 1092–1102. https://doi.org/10.1111/2041-210X.12731

GBIF, 2019. What is GBIF?

Graham, C.H., Hijmans, R.J., 2006. A comparison of methods for mapping species ranges and species richness. Global Ecology and Biogeography 15, 578–587. https://doi.org/10.1111/j.1466-8238.2006.00257.x

Guisan, A., Rahbek, C., 2011. SESAM - a new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages: Predicting spatio-temporal patterns of species assemblages. Journal of Biogeography 38, 1433–1444. https://doi.org/10.1111/j.1365-2699.2011.02550.x

Guisan, A., Thuiller, W., 2005. Predicting species distribution: Offering more than simple habitat models. Ecology Letters 8, 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecological modelling 135, 147–186.

Hirzel, A.H., Helfer, V., Metral, F., 2001. Assessing habitat-suitability models with a virtual species. Ecological Modelling 145, 111–121. https://doi.org/10.1016/S0304-3800(01)00396-9

Hubbell, S.P., 2001. The unified neutral theory of biodiversity and biogeography, Monographs in population biology. Princeton University Press, Princeton.

Hurlbert, A.H., Stegen, J.C., 2014. When should species richness be energy limited, and how would we know? Ecology Letters 17, 401–413. https://doi.org/10.1111/ele.12240

Hutchinson, G.E., 1957. Concluding Remarks. Cold Spring Harbor Symposia on Quantitative Biology 22, 415–427. https://doi.org/10.1101/SQB.1957.022.01.039

Jim'enez-Valverde, A., Lobo, J.M., 2007. Threshold criteria for conversion of probability of species presence to eitheror presenceabsence. Acta Oecologica 31, 361–369. https://doi.org/10.1016/j.actao.2007.02.001

Keil, P., Belmaker, J., Wilson, A.M., Unitt, P., Jetz, W., 2013. Downscaling of species distribution models: A hierarchical approach. Journal of Applied Ecology 82–94. https://doi.org/10.1111/j.2041-210x.2012.00264.x@10.1111/(ISSN)1365-2664.INTERNATIONAL

Laughlin, D.C., Messier, J., 2015. Fitness of multidimensional phenotypes in dynamic adaptive landscapes. Trends in Ecology & Evolution 30, 487–496. https://doi.org/10.1016/j.tree.2015.06.003

Leroy, B., Meynard, C.N., Bellard, C., Courchamp, F., 2016. Virtualspecies, an R package to generate virtual species distributions. Ecography 39, 599–607. https://doi.org/10.1111/ecog.01388

Lessard, J.-P., Belmaker, J., Myers, J.A., Chase, J.M., Rahbek, C., 2012. Inferring local ecological processes amid species pool influences. Trends in Ecology & Evolution 27, 600–607. https://doi.org/10.1016/j.tree.2012.07.006

Lessard, J.-P., Weinstein, B.G., Borregaard, M.K., Marske, K.A., Martin, D.R., McGuire, J.A., Parra, J.L., Rahbek, C., Graham, C.H., 2015. Process-Based Species Pools Reveal the Hidden Signature of Biotic Interactions Amid the Influence of Temperature Filtering. The American Naturalist 187, 75–88. https://doi.org/10.1086/684128

493 Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G., 2005. Selecting thresholds of occurrence in
494 the prediction of species distributions. Ecography 28, 385–393.
495 https://doi.org/10.1111/j.0906-7590.2005.03957.x

496 Liu, C., White, M., Newell, G., 2013. Selecting thresholds for the prediction of species
497 occurrence with presence-only data. Journal of Biogeography 40, 778–789.
498 https://doi.org/10.1111/jbi.12058

499 Mackey, B.G., Lindenmayer, D.B., 2001. Towards a hierarchical framework for modelling the
500 spatial distribution of animals. Journal of Biogeography 28, 1147–1166.
501 https://doi.org/10.1046/j.1365-2699.2001.00626.x

502 Mateo, R.G., Mokany, K., Guisan, A., 2017. Biodiversity Models: What If Unsaturation Is the
503 Rule? Trends in Ecology & Evolution 0. https://doi.org/10.1016/j.tree.2017.05.003

504 Mazel, F., Guilhaumon, F., Mouquet, N., Devictor, V., Gravel, D., Renaud, J., Cianciaruso,
505 M.V., Loyola, R., Diniz-Filho, J.A.F., Mouillot, D., Thuiller, W., 2014. Multifaceted
506 diversityArea relationships reveal global hotspots of mammalian species, trait and lineage
507 diversity. Global Ecology and Biogeography 23, 836–847.
508 https://doi.org/10.1111/geb.12158

509 Mertes, K., Jetz, W., 2018. Disentangling scale dependencies in species environmental niches and
510 distributions. Ecography 41, 1604–1615. https://doi.org/10.1111/ecog.02871

511 Meyer, C.B., 2007. Does Scale Matter in Predicting Species Distributions? Case Study with the
512 Marbled Murrelet. Ecological Applications 17, 1474–1483. https://doi.org/10.1890/06-
513 1410.1

514 Meynard, C.N., Kaplan, D.M., 2013. Using virtual species to study species distributions and
515 model performance. Journal of Biogeography 40, 1–8. https://doi.org/10.1111/jbi.12006

516 Meynard, C.N., Kaplan, D.M., 2012. The effect of a gradual response to the environment on
517 species distribution modeling performance. Ecography 35, 499–509.
518 https://doi.org/10.1111/j.1600-0587.2011.07157.x

519 Meynard, C.N., Leroy, B., Kaplan, D.M., 2019. Testing methods in species distribution
520 modelling using virtual species: What have we learnt and what are we missing?
521 Ecography 0. https://doi.org/10.1111/ecog.04385

522 Munguia, M., Peterson, A.T., S'anchez-Cordero, V., 2008. Dispersal limitation and geographical
523 distributions of mammal species. Journal of Biogeography 35, 1879–1887.
524 https://doi.org/10.1111/j.1365-2699.2008.01921.x

525 Munoz, F., Greni'e, M., Denelle, P., Taudiere, A., Laroche, F., Tucker, C., Violle, C., 2017.
526 Ecolottery: Simulating and assessing community assembly with environmental filtering
527 and neutral dynamics in R. Methods in Ecology and Evolution in press.

528 Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., Kent, J., 2000.
529 Biodiversity hotspots for conservation priorities. Nature 403, 853–858.
530 https://doi.org/10.1038/35002501

531 Newbold, T., Hudson, L.N., Hill, S.L.L., Contu, S., Lysenko, I., Senior, R.A., Börger, L.,
532 Bennett, D.J., Choimes, A., Collen, B., Day, J., De Palma, A., Díaz, S., Echeverria-
533 Londoño, S., Edgar, M.J., Feldman, A., Garon, M., Harrison, M.L.K., Alhusseini, T.,
534 Ingram, D.J., Itescu, Y., Kattge, J., Kemp, V., Kirkpatrick, L., Kleyer, M., Correia, D.L.P.,

Martin, C.D., Meiri, S., Novosolov, M., Pan, Y., Phillips, H.R.P., Purves, D.W., Robinson, A., Simpson, J., Tuck, S.L., Weiher, E., White, H.J., Ewers, R.M., Mace, G.M., Scharlemann, J.P.W., Purvis, A., 2015. Global effects of land use on local terrestrial biodiversity. Nature 520, 45–50. https://doi.org/10.1038/nature14324

O'Brien, E., 1998. Water-energy dynamics, climate, and prediction of woody plant species richness: An interim general model. Journal of Biogeography 25, 379–398. https://doi.org/10.1046/j.1365-2699.1998.252166.x

Oksanen, J., Minchin, P.R., 2002. Continuum theory revisited: What shape are species responses along ecological gradients? Ecological Modelling 157, 119–129. https://doi.org/10.1016/S0304-3800(02)00190-4

Patrick, C.J., Brown, B.L., 2018. Species Pool Functional Diversity Plays a Hidden Role in Generating β-Diversity. The American Naturalist 191, E159–E170. https://doi.org/10.1086/696978

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? Global Ecology and Biogeography 12, 361–371. https://doi.org/10.1046/j.1466-822X.2003.00042.x

Pellissier, L., EspÃndola, A., Pradervand, J.-N., Dubuis, A., Pottier, J., Ferrier, S., Guisan, A., 2013. A probabilistic approach to niche-based community models for spatial forecasts of assemblage properties and their uncertainties. Journal of Biogeography 40, 1939–1946. https://doi.org/10.1111/jbi.12140

555 Pereira, H.M., Belnap, J., Böhm, M., Brummitt, N., Garcia-Moreno, J., Gregory, R., Martin, L.,
556 Peng, C., Proença, V., Schmeller, D., van Swaay, C., 2017. Monitoring Essential
557 Biodiversity Variables at the Species Level, in: Walters, M., Scholes, R.J. (Eds.), The
558 GEO Handbook on Biodiversity Observation Networks. Springer International Publishing,
559 Cham, pp. 79–105. https://doi.org/10.1007/978-3-319-27288-7_4

560 Pereira, H.M., Ferrier, S., Walters, M., Geller, G.N., Jongman, R.H.G., Scholes, R.J., Bruford,
561 M.W., Brummitt, N., Butchart, S.H.M., Cardoso, A.C., Coops, N.C., Dulloo, E., Faith,
562 D.P., Freyhof, J., Gregory, R.D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D.S.,
563 McGeoch, M.A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann,
564 J.P.W., Stuart, S.N., Turak, E., Walpole, M., Wegmann, M., 2013. Essential Biodiversity
565 Variables. Science 339, 277–278. https://doi.org/10.1126/science.1229931

566 Pineda, E., Lobo, J.M., 2009. Assessing the accuracy of species distribution models to predict
567 amphibian species richness patterns. Journal of Animal Ecology 78, 182–190.
568 https://doi.org/10.1111/j.1365-2656.2008.01471.x

569 Pottier, J., Dubuis, A., Pellissier, L., Maiorano, L., Rossier, L., Randin, C.F., Vittoz, P., Guisan,
570 A., 2013. The accuracy of plant assemblage prediction from species distribution models
571 varies along environmental gradients: Climate and species assembly predictions. Global
572 Ecology and Biogeography 22, 52–63. https://doi.org/10.1111/j.1466-8238.2012.00790.x

573 Pouteau, R., Munoz, F., Birnbaum, P., 2019. Disentangling the processes driving tree community
574 assembly in a tropical biodiversity hotspot (New Caledonia). Journal of Biogeography 46,
575 796–806. https://doi.org/10.1111/jbi.13535

576 Pulliam, H.R., 2000. On the relationship between niche and distribution. Ecology letters 3, 349–
577 361.

578 Pulliam, H.R., Danielson, B.J., 1991. Sources, Sinks, and Habitat Selection: A Landscape
579 Perspective on Population Dynamics. The American Naturalist 137, S50–S66.
580 https://doi.org/10.1086/285139

581 R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for
582 Statistical Computing, Vienna, Austria.

583 Scherrer, D., D'Amen, M., Fernandes, R.F., Mateo, R.G., Guisan, A., 2018. How to best
584 threshold and validate stacked species assemblages? Community optimisation might hold
585 the answer. Methods in Ecology and Evolution 0. https://doi.org/10.1111/2041-
586 210X.13041

587 Schmitt, S., Robin Pouteau, Dimitri Justeau, Florian Boissieu, Philippe Birnbaum, Nick Golding,
588 2017. Ssdm: An r package to predict distribution of species richness and composition
589 based on stacked species distribution models. Methods in Ecology and Evolution 8, 1795–
590 1803. https://doi.org/10.1111/2041-210X.12841

591 Stachowicz, J.J., 2001. Mutualism, Facilitation, and the Structure of Ecological
592 CommunitiesPositive interactions play a critical, but underappreciated, role in ecological
593 communities by reducing physical or biotic stresses in existing habitats and by creating
594 new habitats on which many species depend. BioScience 51, 235–246.
595 https://doi.org/10.1641/0006-3568(2001)051[0235:MFATSO]2.0.CO;2

596 Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: A
597 citizen-based bird observation network in the biological sciences. Biological Conservation
598 142, 2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006

599 Svenning, J.-C., Skov, F., 2004. Limited filling of the potential range in European tree species.
600 Ecology Letters 7, 565–573. https://doi.org/10.1111/j.1461-0248.2004.00614.x

601 Tedesco, P.A., Beauchard, O., Bigorne, R., Blanchet, S., Buisson, L., Conti, L., Cornu, J.-F.,
602 Dias, M.S., Grenouillet, G., Hugueny, B., J'ez'equel, C., Leprieur, F., Brosse, S.,
603 Oberdorff, T., 2017. A global database on freshwater fish species occurrence in drainage
604 basins. Scientific Data 4, 170141. https://doi.org/10.1038/sdata.2017.141

605 van der Plas, F., Janzen, T., Ordonez, A., Fokkema, W., Reinders, J., Etienne, R.S., Olff, H.,
606 2015. A new modeling approach estimates the relative importance of different community
607 assembly processes. Ecology 96, 1502–1515. https://doi.org/10.1890/14-0454.1

608 Václavík, T., Meentemeyer, R.K., 2012. Equilibrium or not? Modelling potential distribution of
609 invasive species in different stages of invasion. Diversity and Distributions 18, 73–83.
610 https://doi.org/10.1111/j.1472-4642.2011.00854.x

611 Violle, C., Bonis, A., Plantegenest, M., Cudennec, C., Damgaard, C., Marion, B., Cœur, D.L.,
612 Bouzill'e, J.-B., 2011. Plant functional traits capture species richness variations along a
613 flooding gradient. Oikos 120, 389–398. https://doi.org/10.1111/j.1600-0706.2010.18525.x

614

Figures

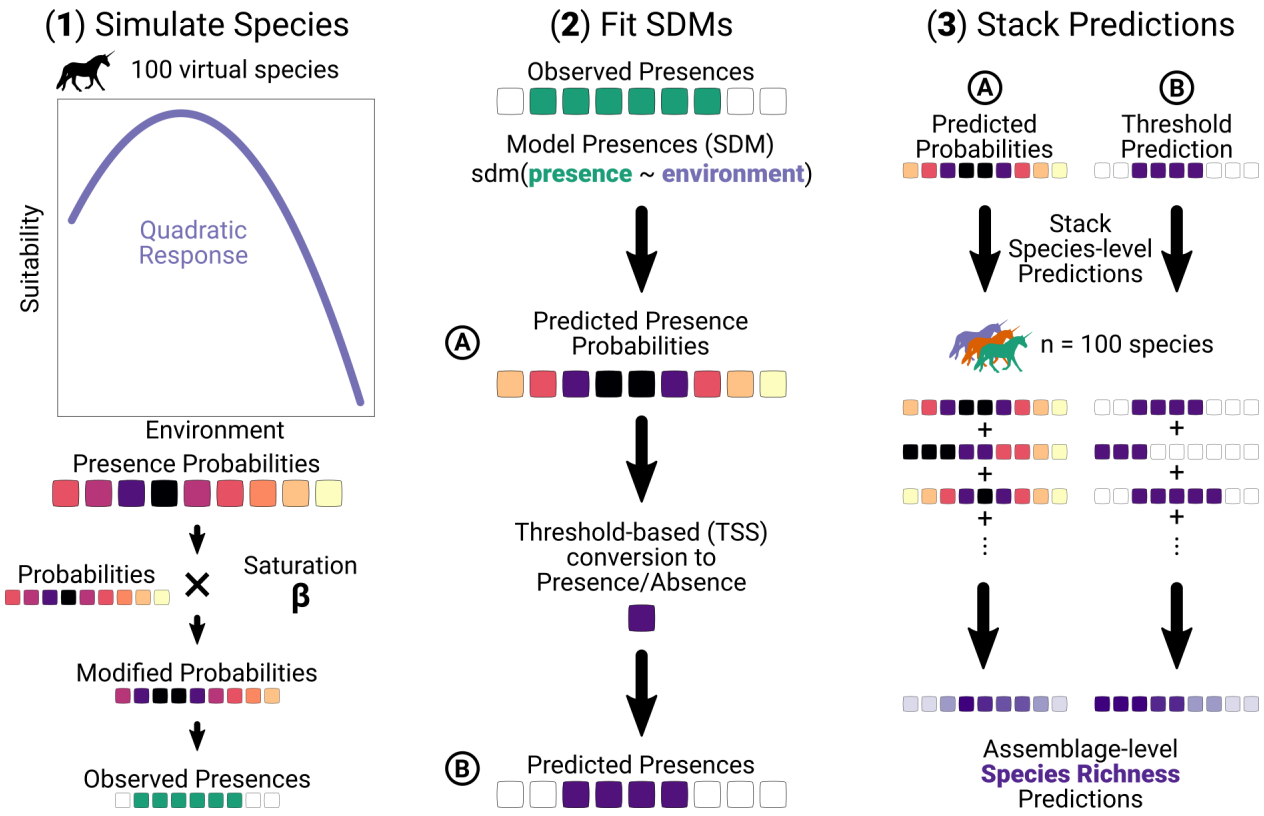


Figure 1: Full Simulation routine. **(1)** We first simulated 100 virtual species with quadratic environmental suitability curves with randomly sampled coefficients. We multiplied each predicted presence probability by the habitat saturation level then use these probabilities to draw realized presences (see Material and Methods for details). Then, using the modified probabilities we drew presences in each assemblage following a binomial distribution. **(2)** We analyzed the realized presences with a binomial Generalized Linear Model (GLM), independently for each species, which provided predicted presence probability of each species in each assemblage **(A)**. We defined a threshold based on True Skill Statistic (see Material and Methods for details, Allouche et al. , 2006). This gave the second set of predictions: **(B)** binary predictions. **(3)** Finally, we summed individual predictions in each assemblage for all the species to get two richness predictions.

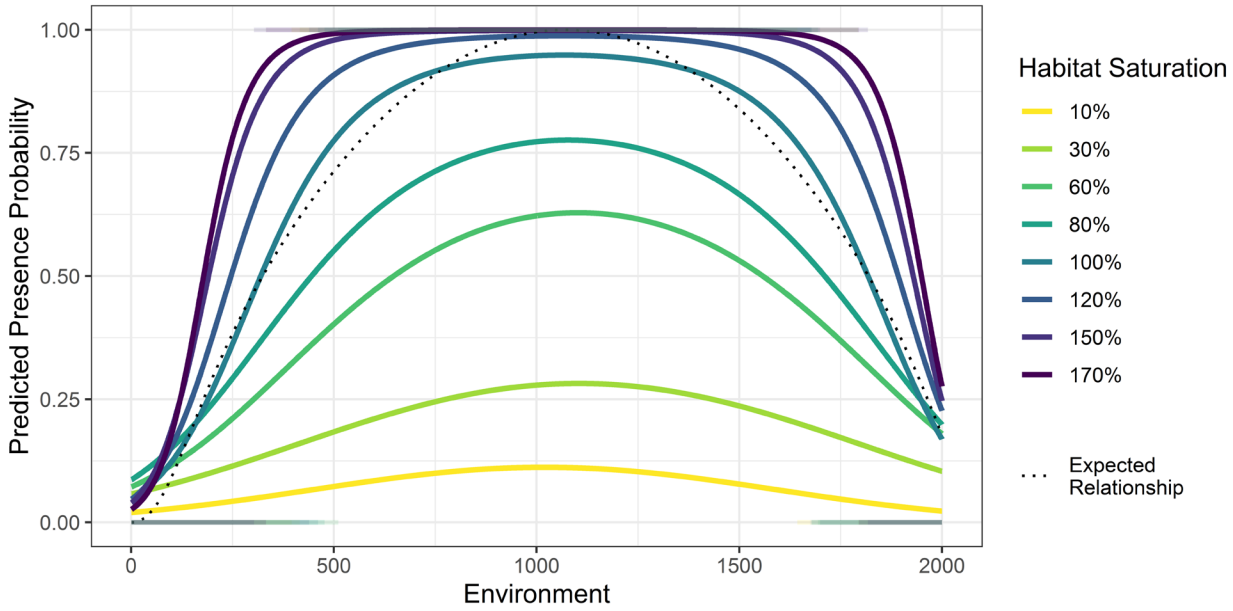
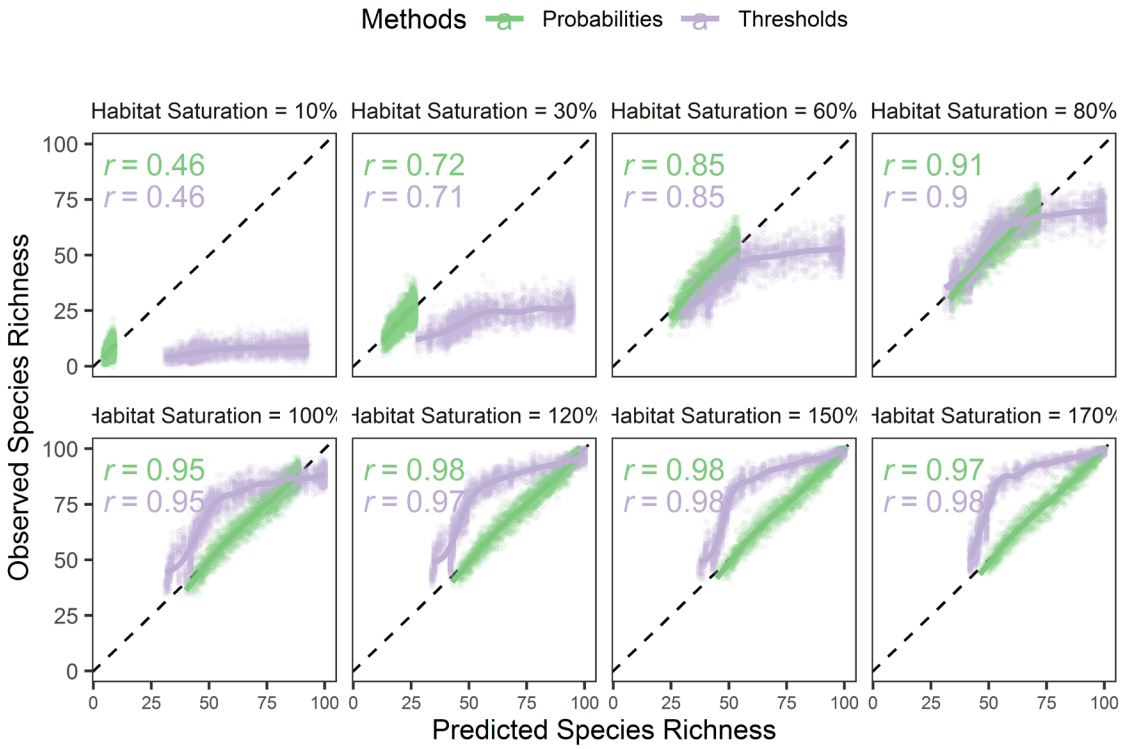
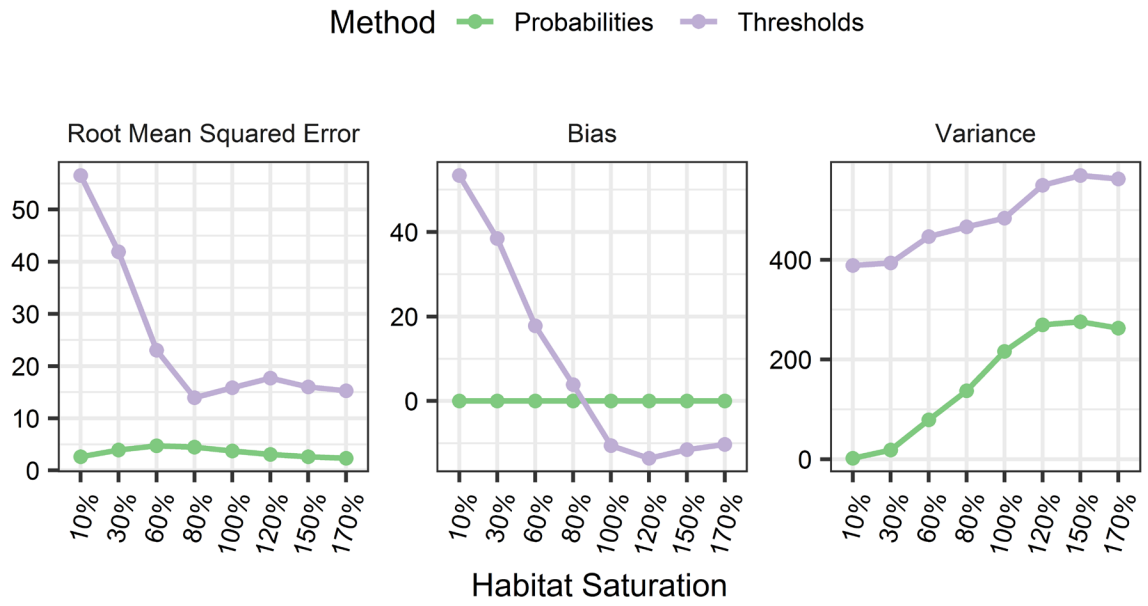


Figure 2: Species expected and predicted presence probability with and without threshold. The solid curves are the predicted presence probabilities by the GLM used to model the presence of species. The dotted curve is the expected relationship given by the parameters of the species. Segments above and below respectively show predicted presences and absences using species-specific threshold.



634

635 Figure 3: Observed vs. predicted richness between two prediction methods as a function of
636 habitat saturation. Each facet shows different species habitat saturation (see Material and
637 Methods). The dashed line is the identity line (y = x), indicating perfect predictions. Green points
638 are probability-based richness predictions; Purple points are threshold-based richness predictions.
639 The corresponding colored lines are cubic splines smoothers trend lines. Spearman correlation
640 coefficients are shown in the top left corner of each facet.



641

642 Figure 4: Prediction accuracy of probability-based and threshold-based richness predictions in
643 function of habitat saturation. Green points and lines: probability-based richness; purple points
644 and lines: threshold-based richness. **(left)** Root Mean Square Error (RMSE) of predicted richness,
645 the average error of richness prediction; **(middle)** Bias, the average difference across all
646 assemblages between predicted and observed richness; **(right)** Variance, the variance of richness
647 predictions across all assemblages.